Survival Analysis: Introduction

Survival Analysis typically focuses on **time to event** data. In the most general sense, it consists of techniques for positive-valued random variables, such as

- time to death
- time to onset (or relapse) of a disease
- length of stay in a hospital
- duration of a strike
- money paid by health insurance
- viral load measurements
- time to finishing a doctoral dissertation!

Kinds of survival studies include:

- clinical trials
- prospective cohort studies
- retrospective cohort studies
- retrospective correlative studies

Typically, survival data are not fully observed, but rather are *censored*.

In this course, we will:

- **describe survival data**
- **compare survival of several groups**
- **explain survival with covariates**
- **design studies with survival endpoints**

Some knowledge of discrete data methods will be useful, since analysis of the "time to event" uses information from the discrete (i.e., binary) outcome of whether the event occurred or not.

Some useful references:

- Collett: *Modelling Survival Data in Medical Research*
- Cox and Oakes: *Analysis of Survival Data*
- Kalbfleisch and Prentice: *The Statistical Analysis of Failure Time Data*
- Lee: *Statistical Methods for Survival Data Analysis*
- Fleming & Harrington: *Counting Processes and Survival Analysis*
- Hosmer & Lemeshow: *Applied Survival Analysis*
- Kleinbaum: *Survival Analysis: A self-learning text*

- Klein & Moeschberger: *Survival Analysis: Techniques for censored and truncated data*
- Cantor: *Extending SAS Survival Analysis Techniques for Medical Research*
- Allison: *Survival Analysis Using the SAS System*
- Jennison & Turnbull: *Group Sequential Methods with Applications to Clinical Trials*
- Ibrahim, Chen, & Sinha: *Bayesian Survival Analysis*

Some Definitions and notation

Failure time random variables are always **non-negative**. That is, if we denote the failure time by T, then $T \geq 0$.

T can either be **discrete** (taking a finite set of values, e.g. $a_1, a_2, \ldots, a_n$) or **continuous** (defined on $(0, \infty)$).

A random variable X is called a **censored failure time random variable** if $X = \min(T, U)$, where U is a non-negative censoring variable.

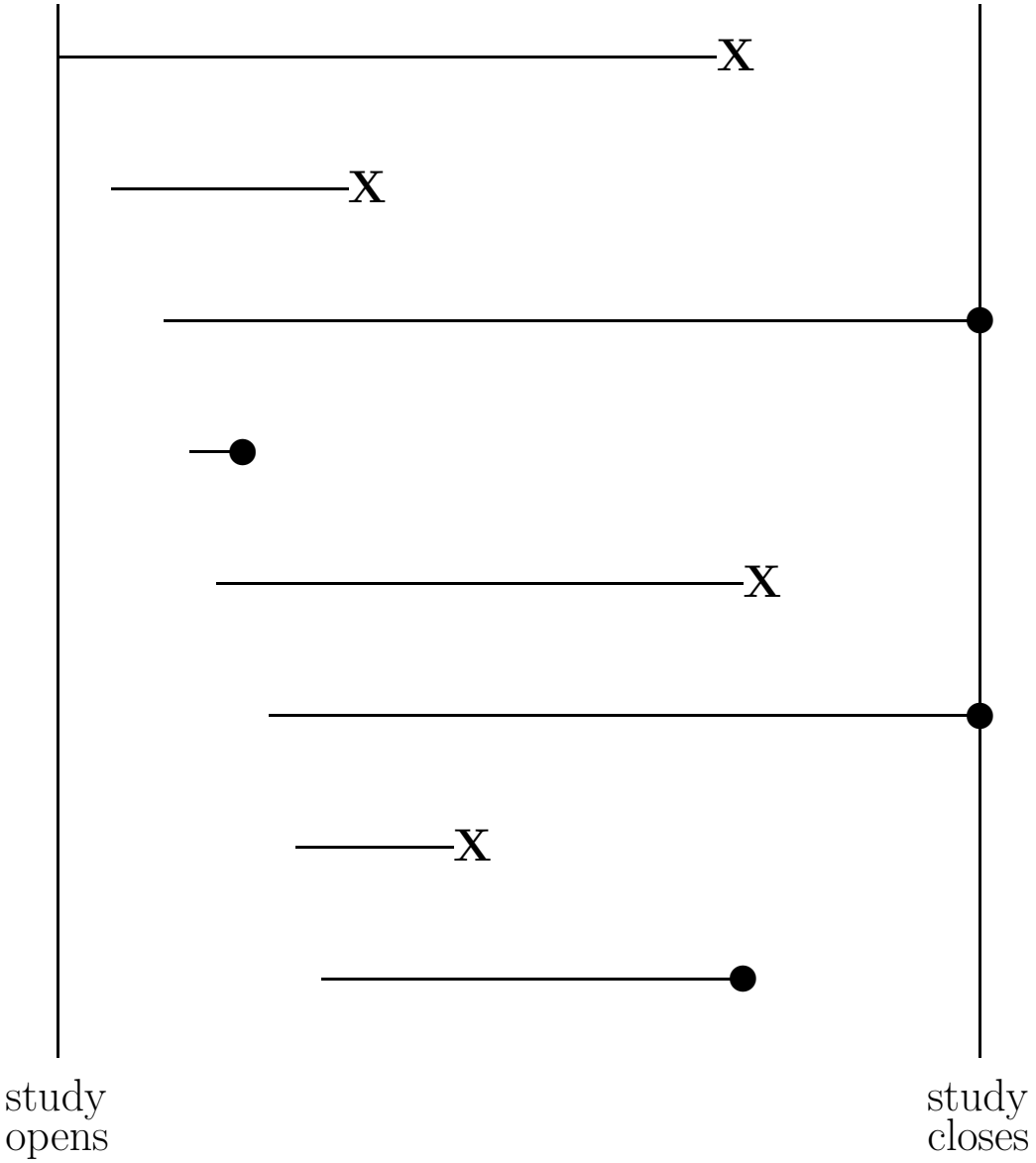
In order to define a failure time random variable, we need:

(1) an unambiguous **time origin**
(e.g. randomization to clinical trial, purchase of car)

(2) a **time scale**
(e.g. real time (days, years), mileage of a car)

(3) definition of the **event**
(e.g. death, need a new car transmission)

Illustration of survival data



● = censored observation

X = event

The illustration of survival data on the previous page shows several features which are typically encountered in analysis of survival data:

- individuals do not all enter the study at the same time
- when the study ends, some individuals still haven't had the event yet
- other individuals drop out or get lost in the middle of the study, and all we know about them is the last time they were still "free" of the event

The first feature is referred to as **"staggered entry"**

The last two features relate to **"censoring"** of the failure time events.

Types of censoring:

- **Right-censoring**:

  only the r.v. $X_i = \min(T_i, U_i)$ is observed due to

  – loss to follow-up

  – drop-out

  – study termination

  We call this right-censoring because the true unobserved event is to the right of our censoring time; i.e., all we know is that the event has not happened at the end of follow-up.

  In addition to observing X_i, we also get to see the **failure indicator**:

$$\delta_i = \begin{cases} 1 & \text{if } T_i \leq U_i \\ 0 & \text{if } T_i > U_i \end{cases}$$

  Some software packages instead assume we have a **censoring indicator**:

$$c_i = \begin{cases} 0 & \text{if } T_i \leq U_i \\ 1 & \text{if } T_i > U_i \end{cases}$$

  Right-censoring is the most common type of censoring assumption we will deal with in survival analysis.

- **Left-censoring**

  Can only observe $Y_i = \max(T_i, U_i)$ and the failure indicators:

$$\delta_i = \begin{cases} 1 & \text{if } U_i \leq T_i \\ 0 & \text{if } U_i > T_i \end{cases}$$

  e.g. (Miller) study of age at which African children learn a task. Some already knew (left-censored), some learned during study (exact), some had not yet learned by end of study (right-censored).

- **Interval-censoring**

  Observe (L_i, R_i) where $T_i \in (L_i, R_i)$

  Ex. 1: Time to prostate cancer, observe longitudinal PSA measurements

  Ex. 2: Time to undetectable viral load in AIDS studies, based on measurements of viral load taken at each clinic visit

  Ex. 3: Detect recurrence of colon cancer after surgery. Follow patients every 3 months after resection of primary tumor.

Independent vs informative censoring

- We say censoring is **independent** (non-informative) if U_i is independent of T_i.
  - **Ex. 1** If U_i is the planned end of the study (say, 2 years after the study opens), then it is usually independent of the event times.
  - **Ex. 2** If U_i is the time that a patient drops out of the study because he/she got much sicker and/or had to discontinue taking the study treatment, then U_i and T_i are probably not independent.

  An individual censored at U should be representative of all subjects who survive to U.

  This means that censoring at U *could* depend on prognostic characteristics measured at baseline, but that among all those with the same baseline characteristics, the probability of censoring prior to or at time U should be the same.

- Censoring is considered **informative** if the distribution of U_i contains any information about the parameters characterizing the distribution of T_i.

Suppose we have a sample of observations on n people:

$$(T_1, U_1), (T_2, U_2), ..., (T_n, U_n)$$

There are three main types of (right) censoring times:

- **Type I:** All the U_i's are the same
  e.g. animal studies, all animals sacrificed after 2 years

- **Type II:** $U_i = T_{(r)}$, the time of the rth failure.
  e.g. animal studies, stop when 4/6 have tumors

- **Type III:** the U_i's are random variables, δ_i's are failure indicators:

$$\delta_i = \begin{cases} 1 & \text{if } T_i \leq U_i \\ 0 & \text{if } T_i > U_i \end{cases}$$

Type I and **Type II** are called *singly* censored data,
Type III is called *randomly* censored (or sometimes *progressively* censored).

Some example datasets:

Example A. Duration of nursing home stay

(Morris et al., *Case Studies in Biometry*, Ch 12)

The National Center for Health Services Research studied 36 for-profit nursing homes to assess the effects of different financial incentives on length of stay. "Treated" nursing homes received higher per diems for Medicaid patients, and bonuses for improving a patient's health and sending them home.

Study included 1601 patients admitted between May 1, 1981 and April 30, 1982.

Variables include:

LOS - Length of stay of a resident (in days)
AGE - Age of a resident
RX - Nursing home assignment (1:bonuses, 0:no bonuses)
GENDER - Gender (1:male, 0:female)
MARRIED - (1: married, 0:not married)
HEALTH - health status (2:second best, 5:worst)
CENSOR - Censoring indicator (1:censored, 0:discharged)

First few lines of data:

37 86 1 0 0 2 0
61 77 1 0 0 4 0

Example B. Fecundability

Women who had recently given birth were asked to recall how long it took them to become pregnant, and whether or not they smoked during that time. The outcome of interest (summarized below) is time to pregnancy (measured in menstrual cycles).

19 subjects were not able to get pregnant after 12 months.

Cycle	Smokers	Non-smokers
1	29	198
2	16	107
3	17	55
4	4	38
5	3	18
6	9	22
7	4	7
8	5	9
9	1	5
10	1	3
11	1	6
12	3	6
12+	7	12

Example C: MAC Prevention Clinical Trial

ACTG 196 was a randomized clinical trial to study the effects of combination regimens on prevention of MAC (*mycobacterium avium complex*), one of the most common opportunistic infections in AIDS patients.

The **treatment regimens** were:

- clarithromycin (new)
- rifabutin (standard)
- clarithromycin plus rifabutin

Other characteristics of trial:

- Patients enrolled between April 1993 and February 1994
- Follow-up ended August 1995
- In February 1994, rifabutin dosage was reduced from 3 pills/day (450mg) to 2 pills/day (300mg) due to concern over **uveitis**[1]

The main intent-to-treat analysis compared the 3 treatment arms without adjusting for this change in dosage.

[1] *Uveitis* is an adverse experience resulting in inflammation of the uveal tract in the eyes (about 3-4% of patients reported uveitis).

Example D: HMO Study of HIV-related Survival

This is hypothetical data used by Hosmer & Lemeshow (described on pages 2-17) containing 100 observations on HIV+ subjects belonging to an Health Maintenance Organization (HMO). The HMO wants to evaluate the survival time of these subjects. In this hypothetical dataset, subjects were enrolled from January 1, 1989 until December 31, 1991. Study follow up then ended on December 31, 1995.

Variables:

ID	Subject ID (1-100)
TIME	Survival time in months
ENTDATE	Entry date
ENDDATE	Date follow-up ended due to death or censoring
CENSOR	Death Indicator (1=death, 0=censor)
AGE	Age of subject in years
DRUG	History of IV Drug Use (0=no,1=yes)

This dataset is used by Hosmer & Lemeshow to motivate some concepts in survival analysis in Chap. 1 of their book.

Example E: UMARU Impact Study (UIS)

This dataset comes from the University of Massachusetts AIDS Research Unit (UMARU) IMPACT Study, a 5-year collaborative research project comprised of two concurrent randomized trials of residential treatment for drug abuse.

(1) **Program A:** Randomized 444 subjects to a 3- or 6-month program of health education and relapse prevention. Clients were taught to recognize "high-risk" situations that are triggers to relapse, and taught skills to cope with these situations without using drugs.

(2) **Program B:** Randomized 184 participants to a 6- or 12-month program with highly structured life-style in a communal living setting.

Variables:

ID	Subject ID (1-628)
AGE	Age in years
BECKTOTA	Beck Depression Score
HERCOC	Heroin or Cocaine Use prior to entry
IVHX	IV Drug use at Admission
NDRUGTX	Number previous drug treatments
RACE	Subject's Race (0=White, 1=Other)
TREAT	Treatment Assignment (0=short, 1=long)
SITE	Treatment Program (0=A,1=B)
LOT	Length of Treatment (days)
TIME	Time to Return to Drug Use (days)
CENSOR	Indicator of Drug Use Relapse (1=yes,0=censored)

Example F: Atlantic Halibut Survival Times

One conservation measure suggested for trawl fishing is a minimum size limit for halibut (32 inches). However, this size limit would only be effective if captured fish below the limit survived until the time of their release. An experiment was conducted to evaluate the survival rates of halibut caught by trawls or longlines, and to assess other factors which might contribute to survival (duration of trawling, maximum depth fished, size of fish, and handling time).

An article by Smith, Waiwood and Neilson, *Survival Analysis for Size Regulation of Atlantic Halibut* in *Case Studies in Biometry* compares parametric survival models to semi-parametric survival models in evaluating this data.

Obs #	Survival *Time* (min)	*Censor*ing Indicator	*Tow Dur*ation (min.)	Diff in *Depth*	*Length* of Fish (cm)	*Handling* Time (min.)	Total *log(catch)* ln(weight)
100	353.0	1	30	15	39	5	5.685
109	111.0	1	100	5	44	29	8.690
113	64.0	0	100	10	53	4	5.323
116	500.0	1	100	10	44	4	5.323
....							

More Definitions and Notation

There are several equivalent ways to characterize the probability distribution of a survival random variable. Some of these are familiar; others are special to survival analysis. We will focus on the following terms:

- The density function $f(t)$
- The survivor function $S(t)$
- The hazard function $\lambda(t)$
- The cumulative hazard function $\Lambda(t)$

- **Density function (or Probability Mass Function) for discrete r.v.'s**
  Suppose that T takes values in $a_1, a_2, \ldots, a_n$.

$$\begin{aligned} f(t) &= Pr(T = t) \\ &= \begin{cases} f_j & \text{if} \quad t = a_j, j = 1, 2, \ldots, n \\ 0 & \text{if} \quad t \neq a_j, j = 1, 2, \ldots, n \end{cases} \end{aligned}$$

- **Density Function for continuous r.v.'s**

$$f(t) = \lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr(t \leq T \leq t + \Delta t)$$

- **Survivorship Function**: $S(t) = P(T \geq t)$.

  In other settings, the cumulative distribution function, $F(t) = P(T \leq t)$, is of interest. In survival analysis, our interest tends to focus on the survival function, $S(t)$.

  For a continuous random variable:

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

  For a discrete random variable:

$$\begin{aligned} S(t) &= \sum_{u \geq t} f(u) \\ &= \sum_{a_j \geq t} f(a_j) \\ &= \sum_{a_j \geq t} f_j \end{aligned}$$

Notes:

- From the definition of $S(t)$ for a continuous variable, $S(t) = 1 - F(t)$ as long as $F(t)$ is absolutely continuous w.r.t the Lebesgue measure. [That is, $F(t)$ has a density function.]
- For a discrete variable, we have to decide what to do if an event occurs exactly at time t; i.e., does that become part of $F(t)$ or $S(t)$?
- To get around this problem, several books define $S(t) = Pr(T > t)$, or else define $F(t) = Pr(T < t)$ (eg. Collett)

- **Hazard Function** $\lambda(t)$

  Sometimes called an *instantaneous failure rate*, the *force of mortality*, or the *age-specific failure rate*.

  – **Continuous random variables:**

$$\begin{aligned}
\lambda(t) &= \lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr(t \leq T < t + \Delta t | T \geq t) \\
&= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \frac{Pr([t \leq T < t + \Delta t] \ \cap \ [T \geq t])}{Pr(T \geq t)} \\
&= \lim_{\Delta t \to 0} \frac{1}{\Delta t} \frac{Pr(t \leq T < t + \Delta t)}{Pr(T \geq t)} \\
&= \frac{f(t)}{S(t)}
\end{aligned}$$

  – **Discrete random variables:**

$$\begin{aligned}
\lambda(a_j) \equiv \lambda_j &= Pr(T = a_j | T \geq a_j) \\
&= \frac{P(T = a_j)}{P(T \geq a_j)} \\
&= \frac{f(a_j)}{S(a_j)} \\
&= \frac{f(t)}{\Sigma_{k:a_k \geq a_j} f(a_k)}
\end{aligned}$$

- **Cumulative Hazard Function** $\Lambda(t)$

  – **Continuous random variables:**

$$\Lambda(t) \;=\; \int_0^t \lambda(u)du$$

  – **Discrete random variables:**

$$\Lambda(t) \;=\; \sum_{k:a_k<t} \lambda_k$$

Relationship between $S(t)$ and $\lambda(t)$

We've already shown that, for a continuous r.v.

$$\lambda(t) = \frac{f(t)}{S(t)}$$

For a left-continuous survivor function $S(t)$, we can show:

$$f(t) \;=\; -S'(t) \qquad \text{or} \quad S'(t) \;=\; -\,f(t)$$

We can use this relationship to show that:

$$\begin{aligned} -\,\frac{d}{dt}[\log S(t)] \;&=\; -\,\left(\frac{1}{S(t)}\right) S'(t) \\ &=\; -\,\frac{-f(t)}{S(t)} \\ &=\; \frac{f(t)}{S(t)} \end{aligned}$$

So another way to write $\lambda(t)$ is as follows:

$$\lambda(t) \;=\; -\frac{d}{dt}[\log S(t)]$$

Relationship between $S(t)$ and $\Lambda(t)$:

- **Continuous case:**

$$\begin{aligned}
\Lambda(t) &= \int_0^t \lambda(u)du \\
&= \int_0^t \frac{f(u)}{S(u)}du \\
&= \int_0^t -\frac{d}{du}\log S(u)du \\
&= -\log S(t) + \log S(0) \\
&\Rightarrow S(t) = e^{-\Lambda(t)}
\end{aligned}$$

- **Discrete case:**

  Suppose that $a_j < t \leq a_{j+1}$. Then

$$\begin{aligned}
S(t) &= P(T \geq a_1, T \geq a_2, \ldots, T \geq a_{j+1}) \\
&= P(T \geq a_1)P(T \geq a_2|T \geq a_1)\cdots P(T \geq a_{j+1}|T \geq a_j) \\
&= (1-\lambda_1)\times\cdots\times(1-\lambda_j) \\
&= \prod_{k:a_k<t}(1-\lambda_k)
\end{aligned}$$

  Cox defines $\Lambda(t) = \Sigma_{k:a_k<t}\log(1-\lambda_k)$ so that $S(t) = e^{-\Lambda(t)}$ in the discrete case, as well.

Measuring Central Tendency in Survival

- **Mean survival** - call this μ

$$\mu = \int_0^\infty u f(u) du \quad \text{for continuous } T$$

$$= \sum_{j=1}^{n} a_j f_j \quad \text{for discrete } T$$

- **Median survival** - call this τ, is defined by

$$S(\tau) = 0.5$$

Similarly, any other percentile could be defined.

In practice, we don't usually hit the median survival at exactly one of the failure times. In this case, the estimated median survival is the *smallest* time τ such that

$$\hat{S}(\tau) \leq 0.5$$

Some hazard shapes seen in applications:

- **increasing**

  e.g. aging after 65

- **decreasing**

  e.g. survival after surgery

- **bathtub**

  e.g. age-specific mortality

- **constant**

  e.g. survival of patients with advanced chronic disease

Estimating the survival or hazard function

We can estimate the survival (or hazard) function in two ways:

- by specifying a parametric model for $\lambda(t)$ based on a particular density function $f(t)$
- by developing an empirical estimate of the survival function (i.e., non-parametric estimation)

If no censoring:
The empirical estimate of the survival function, $\tilde{S}(t)$, is the proportion of individuals with event times greater than t.

With censoring:
If there are censored observations, then $\tilde{S}(t)$ is not a good estimate of the true $S(t)$, so other non-parametric methods must be used to account for censoring (life-table methods, Kaplan-Meier estimator)

Some Parametric Survival Distributions

- The **Exponential** distribution (1 parameter)

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

Check: Does $S(t) = e^{-\Lambda(t)}$?

median: solve $0.5 = S(\tau) = e^{-\lambda\tau}$:

$$\Rightarrow \tau = \frac{-\log(0.5)}{\lambda}$$

mean:

$$\int_0^\infty u\lambda e^{-\lambda u} du = \frac{1}{\lambda}$$

- The **Weibull** distribution (2 parameters)
  Generalizes exponential:

$$
\begin{aligned}
S(t) &= e^{-\lambda t^{\kappa}} \\
f(t) &= \frac{-d}{dt} S(t) = \kappa \lambda t^{\kappa-1} e^{-\lambda t^{\kappa}} \\
\lambda(t) &= \kappa \lambda t^{\kappa-1} \\
\Lambda(t) &= \int_0^t \lambda(u) du = \lambda t^{\kappa}
\end{aligned}
$$

  λ - the *scale* parameter
  κ - the *shape* parameter

  The Weibull distribution is convenient because of its simple form. It includes several hazard shapes:

  $\kappa = 1 \rightarrow$ constant hazard

  $0 < \kappa < 1 \rightarrow$ decreasing hazard

  $\kappa > 1 \rightarrow$ increasing hazard

- **Rayleigh** distribution
  Another 2-parameter generalization of exponential:

$$\lambda(t) = \lambda_0 + \lambda_1 t$$

- **compound exponential**

  $T \sim \exp(\lambda)$, $\lambda \sim g$

$$f(t) = \int_0^\infty \lambda e^{-\lambda t} g(\lambda) d\lambda$$

- **log-normal**, **log-logistic**:
  Possible distributions for T obtained by specifying for $\log T$ any convenient family of distributions, e.g.

  $\log T \sim$ normal (non-monotone hazard)

  $\log T \sim$ logistic

Why use one versus another?

- technical convenience for estimation and inference

- explicit simple forms for $f(t), S(t)$, and $\lambda(t)$.

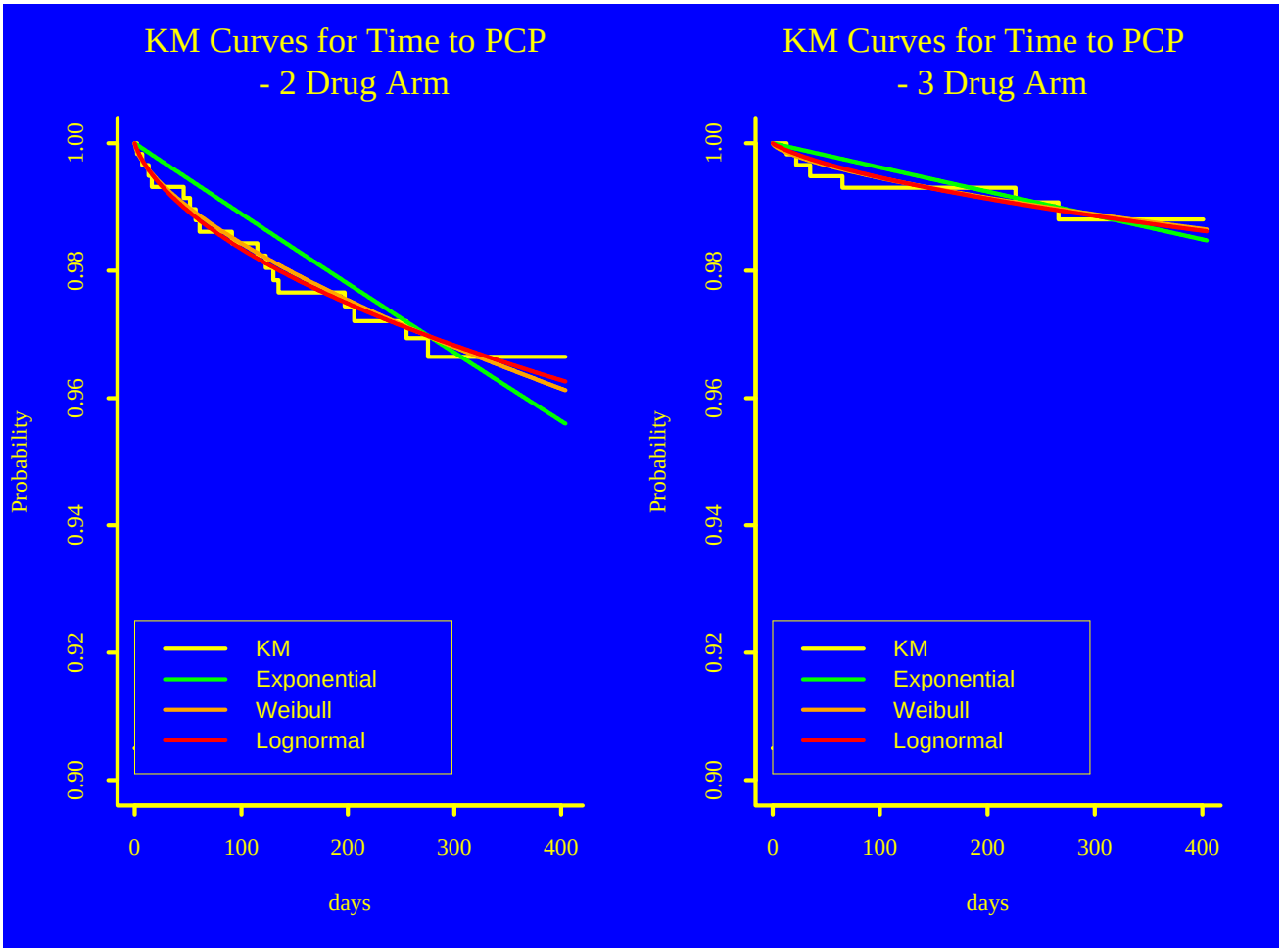
- qualitative shape of hazard function

One can usually distinguish between a one-parameter model (like the exponential) and two-parameter (like Weibull or log-normal) in terms of the adequacy of fit to a dataset.

Without a lot of data, it may be hard to distinguish between the fits of various 2-parameter models (i.e., Weibull vs log-normal)

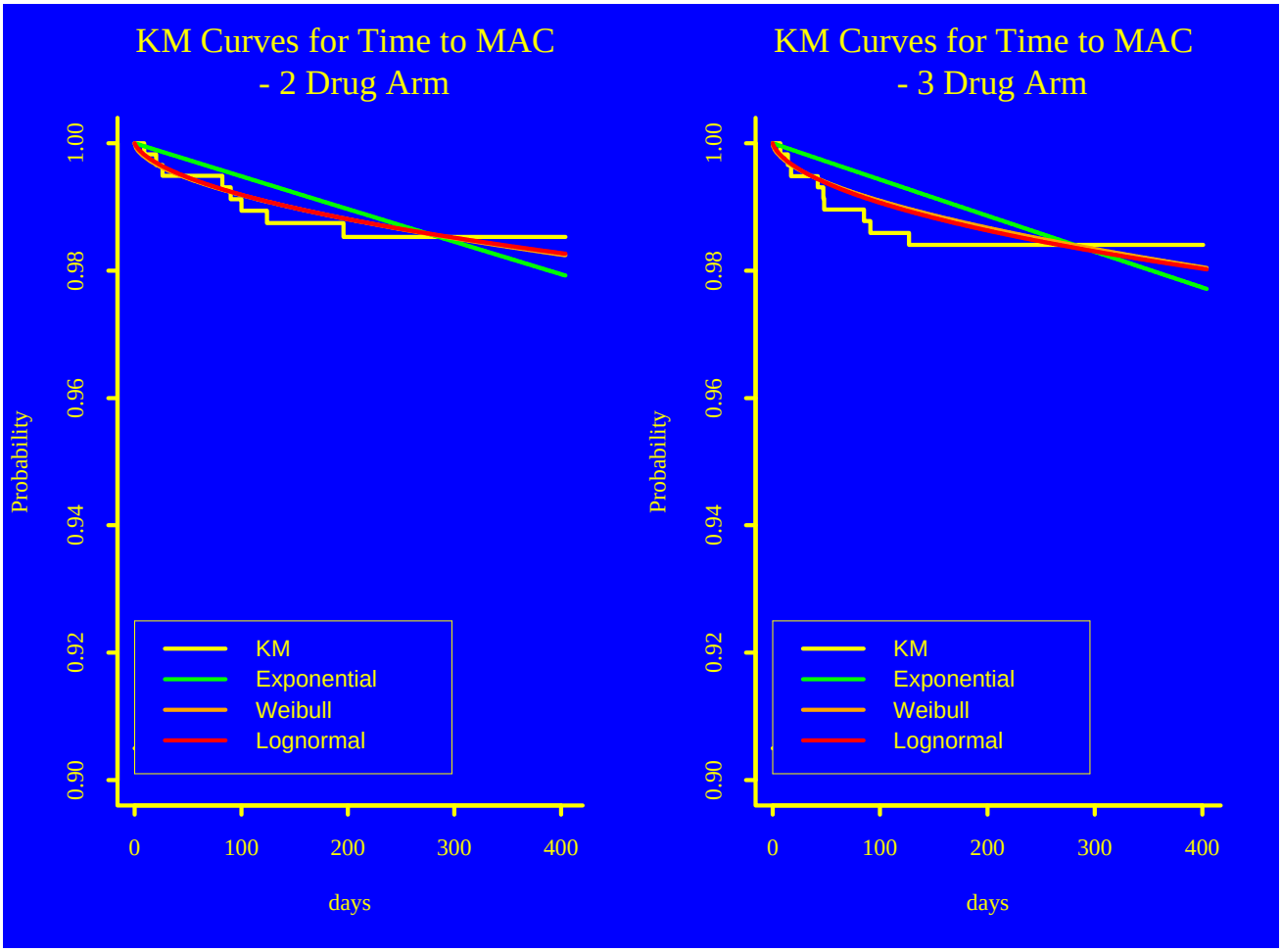
Plots of estimates of $S(t)$

Based on KM, exponential, Weibull, and log-normal
for study of protease inhibitors in AIDS patients
(ACTG 320)



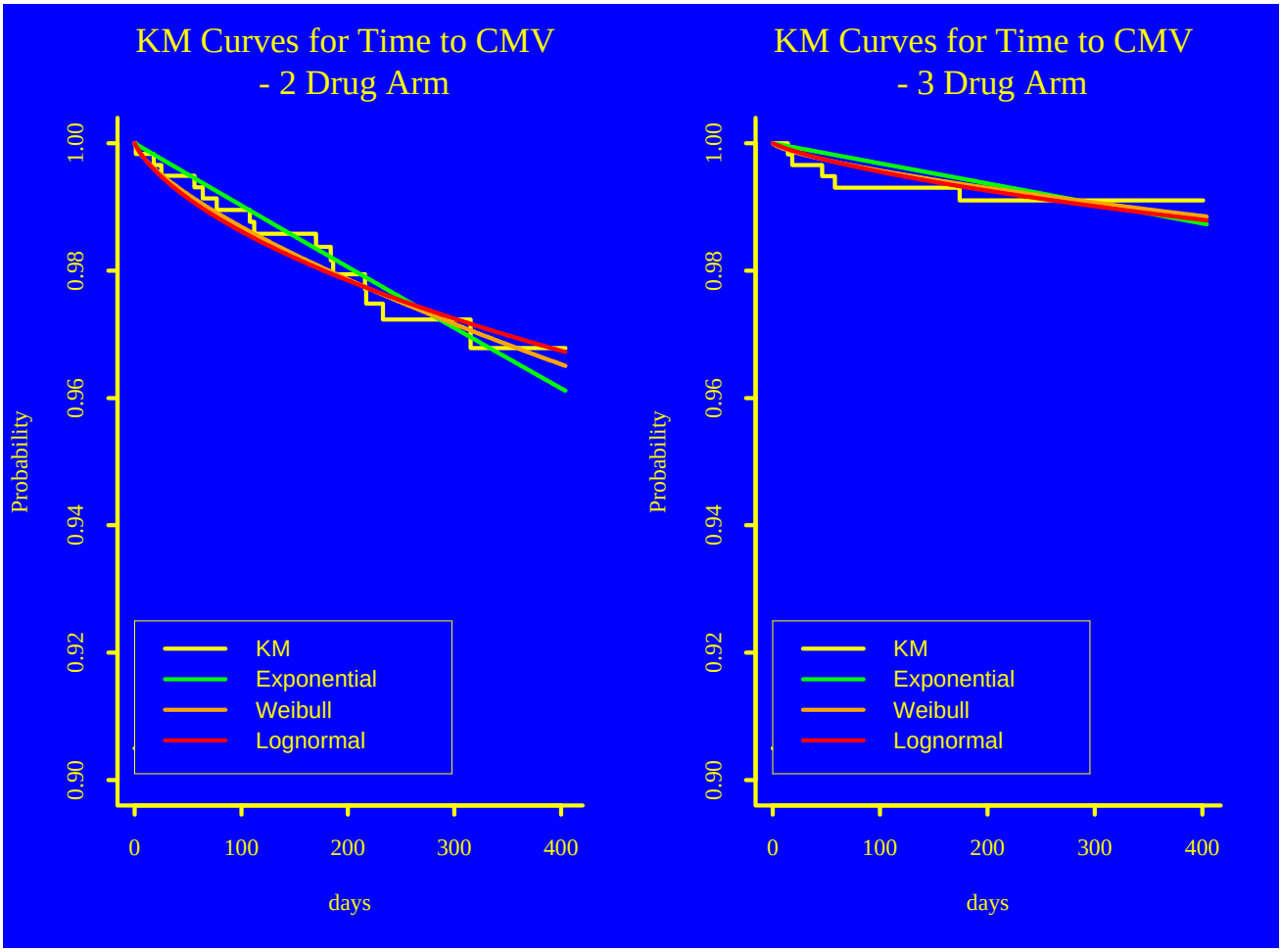
Plots of estimates of $S(t)$

Based on KM, exponential, Weibull, and log-normal
for study of protease inhibitors in AIDS patients
(ACTG 320)



Plots of estimates of $S(t)$

Based on KM, exponential, Weibull, and log-normal
for study of protease inhibitors in AIDS patients
(ACTG 320)



Preview of Coming Attractions

Next we will discuss the most famous non-parametric approach for estimating the survival distribution, called the *Kaplan-Meier estimator*.

To motivate the derivation of this estimator, we will first consider a set of survival times where there is no censoring.

The following are **times to relapse** (weeks) for 21 leukemia patients receiving control treatment (Table 1.1 of Cox & Oakes):

1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

How would we estimate S(10), the probability that an individual survives to time 10 or later?

What about $\tilde{S}(8)$? Is it $\frac{12}{21}$ or $\frac{8}{21}$?

Let's construct a table of $\tilde{S}(t)$:

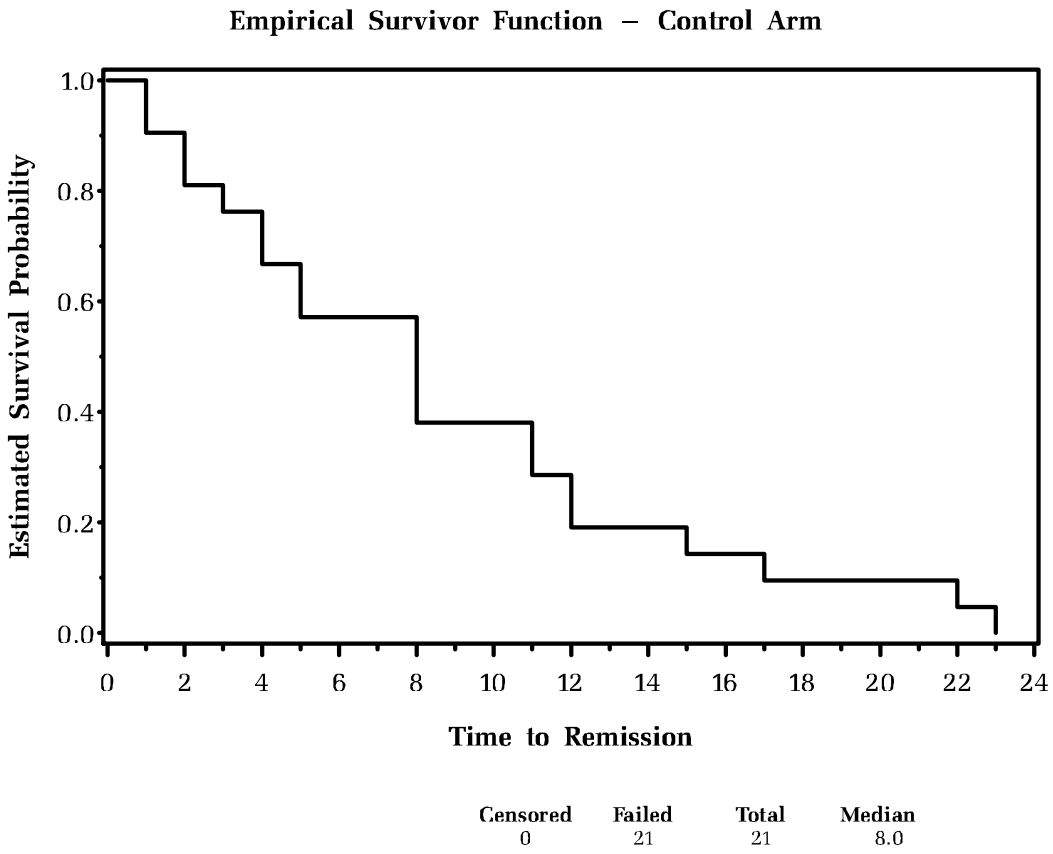
Values of t	$\hat{S}(t)$
$t \leq 1$	21/21=1.000
$1 < t \leq 2$	19/21=0.905
$2 < t \leq 3$	17/21=0.809
$3 < t \leq 4$	
$4 < t \leq 5$	
$5 < t \leq 8$	
$8 < t \leq 11$	
$11 < t \leq 12$	
$12 < t \leq 15$	
$15 < t \leq 17$	
$17 < t \leq 22$	
$22 < t \leq 23$	

Empirical Survival Function:
When there is no censoring, the general formula is:

$$\tilde{S}(t) = \frac{\#\ individuals\ with\ T \geq t}{total\ sample\ size}$$

In most software packages, the survival function is evaluated just after time t, i.e., at t^+. In this case, we only count the individuals with $T > t$.

Example for leukemia data (control arm):



Stata Commands for Survival Estimation

```
.use leukem

.stset remiss status if trt==0       (to keep only untreated patients)
(21 observations deleted)

. sts list
         failure _d:  status
   analysis time _t:  remiss

          Beg.          Net    Survivor     Std.
  Time    Total   Fail   Lost   Function    Error     [95% Conf. Int.]
-------------------------------------------------------------------------
     1       21      2      0     0.9048   0.0641     0.6700    0.9753
     2       19      2      0     0.8095   0.0857     0.5689    0.9239
     3       17      1      0     0.7619   0.0929     0.5194    0.8933
     4       16      2      0     0.6667   0.1029     0.4254    0.8250
     5       14      2      0     0.5714   0.1080     0.3380    0.7492
     8       12      4      0     0.3810   0.1060     0.1831    0.5778
    11        8      2      0     0.2857   0.0986     0.1166    0.4818
    12        6      2      0     0.1905   0.0857     0.0595    0.3774
    15        4      1      0     0.1429   0.0764     0.0357    0.3212
    17        3      1      0     0.0952   0.0641     0.0163    0.2612
    22        2      1      0     0.0476   0.0465     0.0033    0.1970
    23        1      1      0     0.0000        .          .         .
-------------------------------------------------------------------------

.sts graph
```

SAS Commands for Survival Estimation

```
data leuk;
 input t;
cards;
1
1
2
2
3
4
4
5
5
8
8
8
8
11
11
12
12
15
17
22
23
;

proc lifetest data=leuk;
 time t;
run;
```

SAS Output for Survival Estimation

The LIFETEST Procedure

Product-Limit Survival Estimates

t	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000	1.0000	0	0	0	21
1.0000	.	.	.	1	20
1.0000	0.9048	0.0952	0.0641	2	19
2.0000	.	.	.	3	18
2.0000	0.8095	0.1905	0.0857	4	17
3.0000	0.7619	0.2381	0.0929	5	16
4.0000	.	.	.	6	15
4.0000	0.6667	0.3333	0.1029	7	14
5.0000	.	.	.	8	13
5.0000	0.5714	0.4286	0.1080	9	12
8.0000	.	.	.	10	11
8.0000	.	.	.	11	10
8.0000	.	.	.	12	9
8.0000	0.3810	0.6190	0.1060	13	8
11.0000	.	.	.	14	7
11.0000	0.2857	0.7143	0.0986	15	6
12.0000	.	.	.	16	5
12.0000	0.1905	0.8095	0.0857	17	4
15.0000	0.1429	0.8571	0.0764	18	3
17.0000	0.0952	0.9048	0.0641	19	2
22.0000	0.0476	0.9524	0.0465	20	1
23.0000	0	1.0000	0	21	0

SAS Output for Survival Estimation (cont'd)

Summary Statistics for Time Variable t

Quartile Estimates

Percent	Point Estimate	95% Confidence Interval [Lower	Upper)
75	12.0000	8.0000	17.0000
50	8.0000	4.0000	11.0000
25	4.0000	2.0000	8.0000

Mean	Standard Error
8.6667	1.4114

Summary of the Number of Censored and Uncensored Values

Total	Failed	Censored	Percent Censored
21	21	0	0.00

Does anyone have a guess regarding how to calculate the standard error of the estimated survival?

$$\hat{S}(8^+) \;=\; P(T > 8) = \frac{8}{21} \;=\; 0.381$$

(at $t = 8^+$, we count the 4 events at time=8 as already having failed)

$$se[\hat{S}(8^+)] \;=\; 0.106$$

S-Plus Commands for Survival Estimation

```
> t_c(1,1,2,2,3,4,4,5,5,8,8,8,8,11,11,12,12,15,17,22,23)


> surv.fit(t,status=rep(1,21))
 95 percent confidence interval is of type "log"
 time n.risk n.event   survival    std.dev lower 95% CI upper 95% CI
    1     21       2 0.90476190 0.06405645   0.78753505    1.0000000
    2     19       2 0.80952381 0.08568909   0.65785306    0.9961629
    3     17       1 0.76190476 0.09294286   0.59988048    0.9676909
    4     16       2 0.66666667 0.10286890   0.49268063    0.9020944
    5     14       2 0.57142857 0.10798985   0.39454812    0.8276066
    8     12       4 0.38095238 0.10597117   0.22084536    0.6571327
   11      8       2 0.28571429 0.09858079   0.14529127    0.5618552
   12      6       2 0.19047619 0.08568909   0.07887014    0.4600116
   15      4       1 0.14285714 0.07636035   0.05010898    0.4072755
   17      3       1 0.09523810 0.06405645   0.02548583    0.3558956
   22      2       1 0.04761905 0.04647143   0.00703223    0.3224544
   23      1       1 0.00000000         NA           NA           NA
```

Estimating the Survival Function

One-sample nonparametric methods:

We will consider three methods for estimating a survivorship function

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

without resorting to parametric methods:

(1) **Kaplan-Meier**

(2) **Life-table** (Actuarial Estimator)

(3) via the **Cumulative hazard estimator**

(1) The Kaplan-Meier Estimator

The Kaplan-Meier (or KM) estimator is probably the most popular approach. It can be justified from several perspectives:

- product limit estimator
- likelihood justification
- redistribute to the right estimator

We will start with an intuitive motivation based on conditional probabilities, then review some of the other justifications.

Motivation:

First, consider an example where there is no censoring.

The following are times of remission (weeks) for 21 leukemia patients receiving control treatment (Table 1.1 of Cox & Oakes):

1, 1, 2, 2, 3, 4, 4, 5, 5, 8, 8, 8, 8, 11, 11, 12, 12, 15, 17, 22, 23

How would we estimate S(10), the probability that an individual survives to time 10 or later?

What about $\tilde{S}(8)$? Is it $\frac{12}{21}$ or $\frac{8}{21}$?

Let's construct a table of $\tilde{S}(t)$:

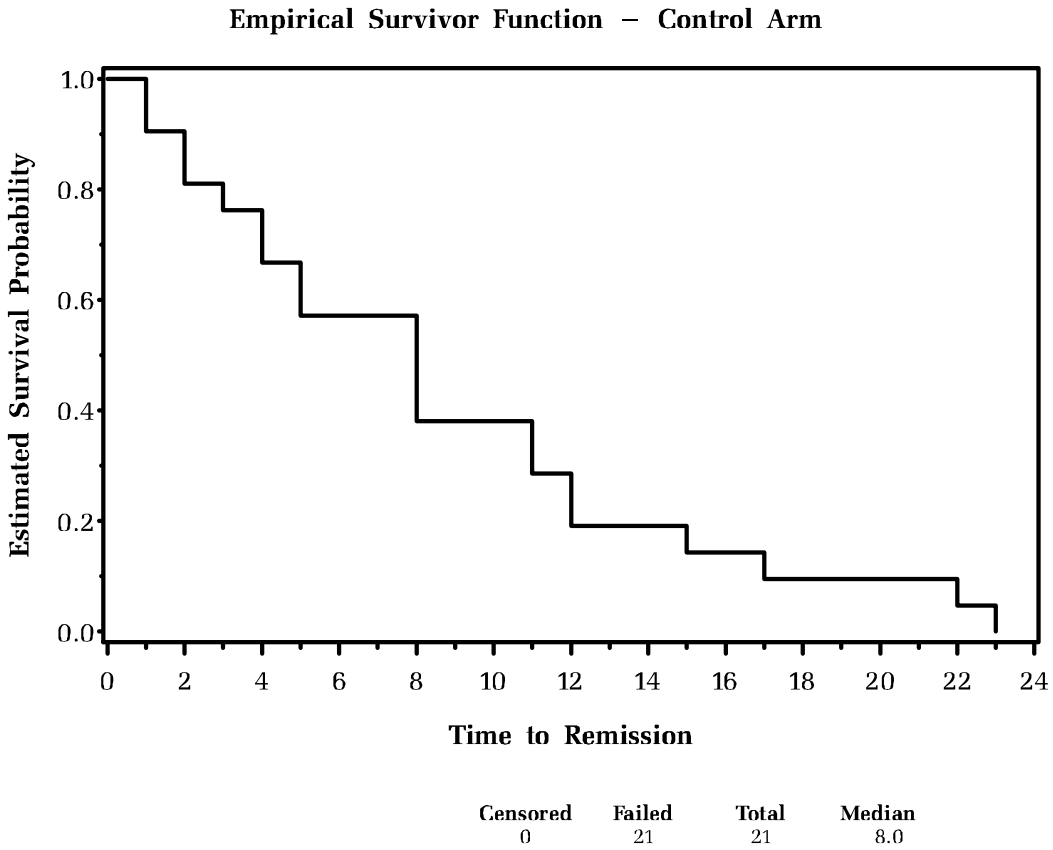
Values of t	$\hat{S}(t)$
$t \leq 1$	21/21=1.000
$1 < t \leq 2$	19/21=0.905
$2 < t \leq 3$	17/21=0.809
$3 < t \leq 4$	
$4 < t \leq 5$	
$5 < t \leq 8$	
$8 < t \leq 11$	
$11 < t \leq 12$	
$12 < t \leq 15$	
$15 < t \leq 17$	
$17 < t \leq 22$	
$22 < t \leq 23$	

Empirical Survival Function:

When there is no censoring, the general formula is:

$$\tilde{S}(t) = \frac{\#\ individuals\ with\ T \geq t}{total\ sample\ size}$$

Example for leukemia data (control arm):



What if there is censoring?

Consider the treated group from Table 1.1 of Cox and Oakes:

$$6^+, 6, 6, 6, 7, 9^+, 10^+, 10, 11^+, 13, 16, 17^+$$
$$19^+, 20^+, 22, 23, 25^+, 32^+, 32^+, 34^+, 35^+$$

[Note: times with $^+$ are right censored]

We know S(6)= 21/21, because everyone survived at least until time 6 or greater. But, we can't say S(7) = 17/21, because we don't know the status of the person who was censored at time 6.

In a 1958 paper in the *Journal of the American Statistical Association*, Kaplan and Meier proposed a way to nonparametrically estimate S(t), even in the presence of censoring. The method is based on the ideas of **conditional probability**.

A quick review of conditional probability:

Conditional Probability: Suppose A and B are two events. Then,

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

Multiplication law of probability: can be obtained from the above relationship, by multiplying both sides by $P(B)$:

$$P(A \cap B) \;=\; P(A|B)\, P(B)$$

Extension to more than 2 events:

Suppose $A_1, A_2 ... A_k$ are k different events. Then, the probability of all k events happening together can be written as a product of conditional probabilities:

$$\begin{aligned} P(A_1 \cap A_2 ... \cap A_k) \;=\; & P(A_k|A_{k-1} \cap ... \cap A_1) \times \\ & \times P(A_{k-1}|A_{k-2} \cap ... \cap A_1) \\ & ... \\ & \times P(A_2|A_1) \\ & \times P(A_1) \end{aligned}$$

Now, let's apply these ideas to estimate $S(t)$:

Suppose $a_k < t \leq a_{k+1}$. Then

$$\begin{aligned}
S(t) &= P(T \geq a_{k+1}) \\
&= P(T \geq a_1, T \geq a_2, \ldots, T \geq a_{k+1}) \\
&= P(T \geq a_1) \times \prod_{j=1}^{k} P(T \geq a_{j+1} | T \geq a_j) \\
&= \prod_{j=1}^{k} [1 - P(T = a_j | T \geq a_j)] \\
&= \prod_{j=1}^{k} [1 - \lambda_j] \\
\text{so} \quad \hat{S}(t) &\cong \prod_{j=1}^{k} \left(1 - \frac{d_j}{r_j}\right) \\
&= \prod_{j:a_j<t} \left(1 - \frac{d_j}{r_j}\right)
\end{aligned}$$

d_j is the number of deaths at a_j
r_j is the number at risk at a_j

Intuition behind the Kaplan-Meier Estimator

Think of dividing the observed timespan of the study into a series of fine intervals so that there is a separate interval for each time of death or censoring:

Using the law of conditional probability,

$$Pr(T \geq t) \;=\; \prod_j Pr(\text{survive } j\text{-th interval } I_j \mid \text{survived to start of } I_j)$$

where the product is taken over all the intervals including or preceding time t.

4 possibilities for each interval:

(1) **No events (death or censoring)** - conditional probability of surviving the interval is 1

(2) **Censoring** - assume they survive to the end of the interval, so that the conditional probability of surviving the interval is 1

(3) **Death, but no censoring** - conditional probability of *not* surviving the interval is # deaths (d) divided by # 'at risk' (r) at the beginning of the interval. So the conditional probability of surviving the interval is $1-(d/r)$.

(4) **Tied deaths and censoring** - assume censorings last to the end of the interval, so that conditional probability of surviving the interval is still $1-(d/r)$

General Formula for jth interval:

It turns out we can write a general formula for the conditional probability of surviving the j-th interval that holds for all 4 cases:

$$1-\frac{d_j}{r_j}$$

We could use the same approach by grouping the event times into intervals (say, one interval for each month), and then counting up the number of deaths (events) in each to estimate the probability of surviving the interval (this is called the *lifetable estimate*).

However, the assumption that those censored last until the end of the interval wouldn't be quite accurate, so we would end up with a cruder approximation.

As the intervals get finer and finer, the approximations made in estimating the probabilities of getting through each interval become smaller and smaller, so that the estimator converges to the true $S(t)$.

This intuition clarifies why an alternative name for the KM is the product limit estimator.

The Kaplan-Meier estimator of the survivorship function (or survival probability) $S(t) = Pr(T \geq t)$ **is:**

$$\begin{aligned} \hat{S}(t) &= \Pi_{j:\tau_j<t} \frac{r_j - d_j}{r_j} \\ &= \Pi_{j:\tau_j<t} \left(1 - \frac{d_j}{r_j}\right) \end{aligned}$$

where

- $\tau_1, ...\tau_K$ is the set of K distinct death times observed in the sample
- d_j is the number of deaths at τ_j
- r_j is the number of individuals "at risk" right before the j-th death time (everyone dead or censored at or after that time).
- c_j is the number of censored observations between the j-th and $(j+1)$-st death times. Censorings tied at τ_j are included in c_j

Note: two useful formulas are:

$$\begin{aligned} (1) \quad r_j &= r_{j-1} - d_{j-1} - c_{j-1} \\ (2) \quad r_j &= \sum_{l \geq j} (c_l + d_l) \end{aligned}$$

Calculating the KM - Cox and Oakes example

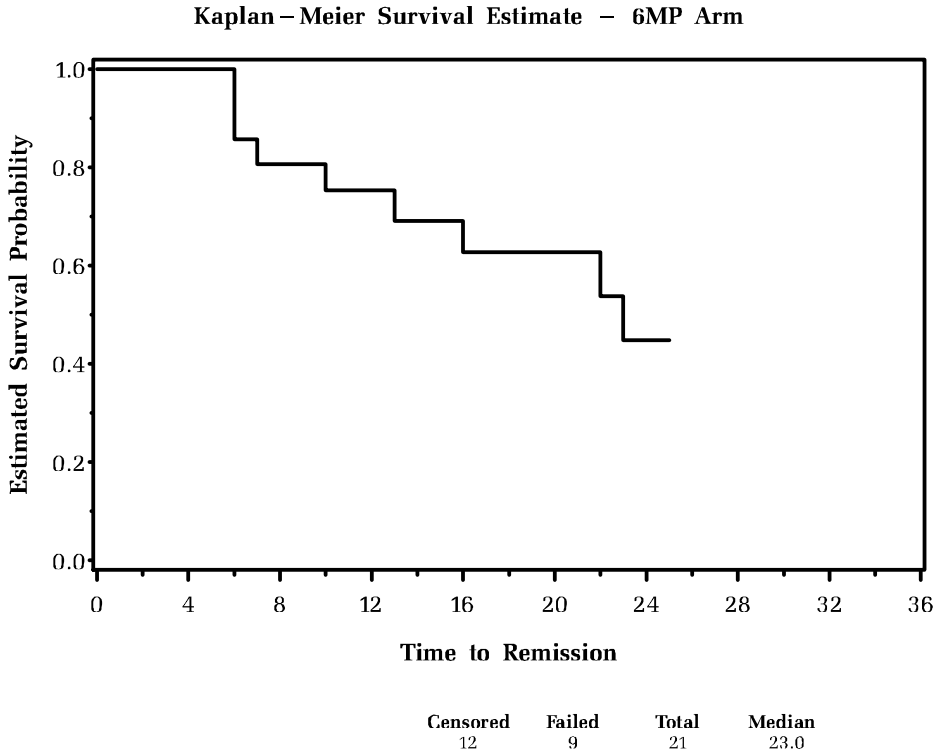
Make a table with a row for every death or censoring time:

τ_j	d_j	c_j	r_j	$1-(d_j/r_j)$	$\hat{S}(\tau_j^+)$
6	3	1	21	$\frac{18}{21}=0.857$	
7	1	0	17		
9	0	1	16		
10					
11					
13					
16					
17					
19					
20					
22					
23					

Note that:

- $\hat{S}(t^+)$ only changes at death (failure) times
- $\hat{S}(t^+)$ is 1 up to the first death time
- $\hat{S}(t^+)$ only goes to 0 if the last event is a death

KM plot for treated leukemia patients



Note: most statistical software packages summarize the KM survival function at τ_j^+, i.e., *just after* the time of the j-th failure.

In other words, they provide $\hat{S}(\tau_j^+)$.

When there is no censoring, the empirical survival estimate would then be:

$$\tilde{S}(t^+) = \frac{\#\ individuals\ with\ T > t}{total\ sample\ size}$$

Output from STATA KM Estimator:

```
    failure time:  weeks
  failure/censor:  remiss
```

Time	Beg. Total	Fail	Net Lost	Survivor Function	Std. Error	[95% Conf.	Int.]
6	21	3	1	0.8571	0.0764	0.6197	0.9516
7	17	1	0	0.8067	0.0869	0.5631	0.9228
9	16	0	1	0.8067	0.0869	0.5631	0.9228
10	15	1	1	0.7529	0.0963	0.5032	0.8894
11	13	0	1	0.7529	0.0963	0.5032	0.8894
13	12	1	0	0.6902	0.1068	0.4316	0.8491
16	11	1	0	0.6275	0.1141	0.3675	0.8049
17	10	0	1	0.6275	0.1141	0.3675	0.8049
19	9	0	1	0.6275	0.1141	0.3675	0.8049
20	8	0	1	0.6275	0.1141	0.3675	0.8049
22	7	1	0	0.5378	0.1282	0.2678	0.7468
23	6	1	0	0.4482	0.1346	0.1881	0.6801
25	5	0	1	0.4482	0.1346	0.1881	0.6801
32	4	0	2	0.4482	0.1346	0.1881	0.6801
34	2	0	1	0.4482	0.1346	0.1881	0.6801
35	1	0	1	0.4482	0.1346	0.1881	0.6801

Two Other Justifications for KM Estimator

I. Likelihood-based derivation (Cox and Oakes)

For a discrete failure time variable, define:

d_j	number of failures at a_j
r_j	number of individuals at risk at a_j (including those censored at a_j).
λ_j	Pr(death) in j-th interval (conditional on survival to start of interval)

The likelihood is that of g independent binomials:

$$L(\boldsymbol{\lambda}) \;=\; \prod_{j=1}^{g} \lambda_j^{d_j}(1-\lambda_j)^{r_j-d_j}$$

Therefore, the **maximum likelihood estimator** of λ_j is:

$$\hat{\lambda}_j = d_j/r_j$$

Now we plug in the MLE's of λ to estimate S(t):

$$\hat{S}(t) \;=\; \prod_{j:a_j<t} (1-\hat{\lambda}_j)$$

$$= \prod_{j:a_j<t} \left(1-\frac{d_j}{r_j}\right)$$

II. Redistribute to the right justification

(Efron, 1967)

In the absence of censoring, $\hat{S}(t)$ is just the proportion of individuals with $T \geq t$. The idea behind Efron's approach is to spread the contributions of censored observations out over all the possible times to their right.

Algorithm:

- Step (1): arrange the n observed times (deaths or censorings) in increasing order. If there are ties, put censored after deaths.
- Step (2): Assign weight $(1/n)$ to each time.
- Step (3): Moving from left to right, each time you encounter a censored observation, distribute its mass to all times to its right.
- Step (4): Calculate $\hat{S}_j$ by subtracting the final weight for time j from $\hat{S}_{j-1}$

Example of "redistribute to the right" algorithm

Consider the following event times:

$$2,\ \ 2.5+,\ \ 3,\ \ 3,\ \ 4,\ \ 4.5+,\ \ 5,\ \ 6,\ \ 7$$

The algorithm goes as follows:

(Step 1) Times	Step 2	Step 3a	Step 3b	(Step 4) $\hat{S}(\tau_j)$
2	1/9=0.11			0.889
2.5^+	1/9=0.11	0		0.889
3	2/9=0.22	0.25		0.635
4	1/9=0.11	0.13		0.508
4.5^+	1/9=0.11	0.13	0	0.508
5	1/9=0.11	0.13	0.17	0.339
6	1/9=0.11	0.13	0.17	0.169
7	1/9=0.11	0.13	0.17	0.000

This comes out the same as the product limit approach.

Properties of the KM estimator

In the case of no censoring:

$$\hat{S}(t) = \tilde{S}(t) = \frac{\#\ deaths\ at\ t\ or\ greater}{n}$$

where n is the number of individuals in the study.

This is just like an estimated probability from a binomial distribution, so we have:

$$\hat{S}(t) \simeq \mathcal{N}\left(S(t),\ S(t)[1 - S(t)]/n\right)$$

How does censoring affect this?

- $\hat{S}(t)$ is still approximately normal
- The mean of $\hat{S}(t)$ converges to the true S(t)
- The variance is a bit more complicated (since the denominator n includes some censored observations).

Once we get the variance, then we can construct (pointwise) $(1 - \alpha)\%$ confidence intervals (NOT bands) about $\hat{S}(t)$:

$$\hat{S}(t) \pm z_{1-\alpha/2}\, se[\hat{S}(t)]$$

Greenwood's formula (Collett 2.1.3)

We can think of the KM estimator as

$$\hat{S}(t) = \prod_{j:\tau_j<t} (1 - \hat{\lambda}_j)$$

where $\hat{\lambda}_j = d_j/r_j$.

Since the $\hat{\lambda}_j$'s are just binomial proportions, we can apply standard likelihood theory to show that each $\hat{\lambda}_j$ is approximately normal, with mean the true λ_j, and

$$var(\hat{\lambda}_j) \approx \frac{\hat{\lambda}_j(1 - \hat{\lambda}_j)}{r_j}$$

Also, the $\hat{\lambda}_j$'s are independent in large enough samples.

Since $\hat{S}(t)$ is a function of the λ_j's, we can estimate its variance using the **delta method**:

> **Delta method**: If Y is normal with mean μ and variance σ^2, then $g(Y)$ is approximately normally distributed with mean $g(\mu)$ and variance $[g'(\mu)]^2\sigma^2$.

Two specific examples of the delta method:

(A) $Z = log(Y)$

$$\text{then} \quad Z \sim N\left[log(\mu), \left(\frac{1}{\mu}\right)^2 \sigma^2\right]$$

(B) $Z = exp(Y)$

$$\text{then} \quad Z \sim N\left[e^{\mu}, [e^{\mu}]^2 \sigma^2\right]$$

The examples above use the following results from calculus:

$$\frac{d}{dx} \log u = \frac{1}{u} \left(\frac{du}{dx}\right)$$

$$\frac{d}{dx} e^u = e^u \left(\frac{du}{dx}\right)$$

Greenwood's formula (continued)

Instead of dealing with $\hat{S}(t)$ directly, we will look at its log:

$$\log[\hat{S}(t)] = \sum_{j:\tau_j<t} \log(1-\hat{\lambda}_j)$$

Thus, by approximate independence of the $\hat{\lambda}_j$'s,

$$\begin{aligned} var(\log[\hat{S}(t)]) &= \sum_{j:\tau_j<t} var[\log(1-\hat{\lambda}_j)] \\ \text{by (A)} \qquad &= \sum_{j:\tau_j<t} \left(\frac{1}{1-\hat{\lambda}_j}\right)^2 var(\hat{\lambda}_j) \\ &= \sum_{j:\tau_j<t} \left(\frac{1}{1-\hat{\lambda}_j}\right)^2 \hat{\lambda}_j(1-\hat{\lambda}_j)/r_j \\ &= \sum_{j:\tau_j<t} \frac{\hat{\lambda}_j}{(1-\hat{\lambda}_j)r_j} \\ &= \sum_{j:\tau_j<t} \frac{d_j}{(r_j-d_j)r_j} \end{aligned}$$

Now, $\hat{S}(t) = \exp[\log[\hat{S}(t)]]$. Thus by (B),

$$var(\hat{S}(t)) = [\hat{S}(t)]^2 var\left[\log[\hat{S}(t)]\right]$$

Greenwood's Formula:

$$var(\hat{S}(t)) = [\hat{S}(t)]^2 \, \Sigma_{j:\tau_j<t} \, \frac{d_j}{(r_j-d_j)r_j}$$

Back to confidence intervals

For a 95% confidence interval, we could use

$$\hat{S}(t) \pm z_{1-\alpha/2}\, se[\hat{S}(t)]$$

where $se[\hat{S}(t)]$ is calculated using Greenwood's formula.

Problem: This approach can yield values > 1 or < 0.

Better approach: Get a 95% confidence interval for

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

Since this quantity is unrestricted, the confidence interval will be in the proper range when we transform back.

To see why this works, note the following:

- Since $\hat{S}(t)$ is an estimated probability

$$0 \leq \hat{S}(t) \leq 1$$

- Taking the log of $\hat{S}(t)$ has bounds:

$$-\infty \leq \log[\hat{S}(t)] \leq 0$$

- Taking the opposite:

$$0 \leq -\log[\hat{S}(t)] \leq \infty$$

- Taking the log again:

$$-\infty \leq \log\left[-\log[\hat{S}(t)]\right] \leq \infty$$

To transform back, reverse steps with $S(t) = \exp(-\exp(L(t))$

Log-log Approach for Confidence Intervals:

(1) Define $L(t) = \log(-\log(S(t)))$

(2) Form a 95% confidence interval for $L(t)$ based on $\hat{L}(t)$, yielding $[\hat{L}(t) - A, \hat{L}(t) + A]$

(3) Since $S(t) = \exp(-\exp(L(t))$, the confidence bounds for the 95% CI on $S(t)$ are:

$$[exp(-e^{(\hat{L}(t)+A)}), exp(-e^{(\hat{L}(t)-A)})]$$

(note that the upper and lower bounds switch)

(4) Substituting $\hat{L}(t) = \log(-\log(\hat{S}(t)))$ back into the above bounds, we get confidence bounds of

$$([\hat{S}(t)]^{e^A}, [\hat{S}(t)]^{e^{-A}})$$

What is A?

- A is $1.96\, se(\hat{L}(t))$
- To calculate this, we need to calculate

$$var(\hat{L}(t)) \;=\; var\left[\log(-\log(\hat{S}(t)))\right]$$

- From our previous calculations, we know

$$var(\log[\hat{S}(t)]) \;=\; \sum_{j:\tau_j<t} \frac{d_j}{(r_j - d_j)r_j}$$

- Applying the delta method as in example (A), we get:

$$\begin{aligned} var(\hat{L}(t)) \;&=\; var(\log(-\log[\hat{S}(t)])) \\ &=\; \frac{1}{[\log\ \hat{S}(t)]^2} \sum_{j:\tau_j<t} \frac{d_j}{(r_j - d_j)r_j} \end{aligned}$$

- We take the square root of the above to get $se(\hat{L}(t))$, and then form the confidence intervals as:

$$\hat{S}(t)^{e^{\pm 1.96\, se(\hat{L}(t))}}$$

- This is the approach that Stata uses. Splus gives an option to calculate these bounds (use `conf.type=''log-log''` in `surv.fit`).

Summary of Confidence Intervals on $S(t)$

- Calculate $\hat{S}(t) \pm 1.96\, se[\hat{S}(t)]$ where $se[\hat{S}(t)]$ is calculated using Greenwood's formula, and replace negative lower bounds by 0 and upper bounds greater than 1 by 1.
  - Recommended by Collett
  - This is the default using SAS
  - not very satisfactory

- Use a log transformation to stabilize the variance and allow for non-symmetric confidence intervals. This is what is normally done for the confidence interval of an estimated odds ratio.
  - Use $var[\log(\hat{S}(t))] = \Sigma_{j:\tau_j<t} \frac{d_j}{(r_j - d_j)r_j}$ already calculated as part of Greenwood's formula
  - This is the default in Splus

- Use the log-log transformation just described
  - Somewhat complicated, but always yields proper bounds
  - This is the default in Stata.

Software for Kaplan-Meier Curves

- Stata - stset and sts commands
- SAS - PROC LIFETEST
- Splus - surv.fit(time,status)

Defaults for Confidence Interval Calculations

- Stata - "log-log" $\Rightarrow \hat{L}(t) \pm 1.96\, se[\hat{L}(t)]$
  where $L(t) = \log[-\log(S(t))]$
- SAS - "plain" $\Rightarrow \hat{S}(t) \pm 1.96\, se[\hat{S}(t)]$
- Splus - "log" $\Rightarrow \log S(t) \pm 1.96\, se[\log(\hat{S}(t))]$

  but Splus will also give either of the other two options if you request them.

Stata Commands

Create a file called "leukemia.dat" with the raw data, with a column for treatment, weeks to relapse (i.e., duration of remission), and relapse status:

```
.infile trt remiss status using leukemia.dat

.stset remiss status          (sets up a failure time dataset,
                               with failtime status in that order,
                               type help stset to get details)

.sts list                    (estimated S(t), se[S(t)], and 95% CI)

.sts graph, saving(kmtrt)    (creates a Kaplan-Meier plot, and
                              saves the plot in file kmtrt.gph,
                              type ‘‘help gphdot’’ to get some
                              printing instructions)

.graph using kmtrt           (redisplays the graph at any later time)
```

If the dataset has already been created and loaded into Stata, then you can substitute the following commands for initializing the data:

```
.use leukem                  (finds Stata dataset leukem.dta)

.describe                    (provides a description of the dataset)

.stset remiss status         (declares data to be failure type)

.stdes                       (gives a description of the survival dataset)
```

STATA Output for Treated Leukemia Patients:

```
.use leukem

.stset remiss status if trt==1

.sts list

    failure time:  remiss
  failure/censor:  status
```

Time	Beg. Total	Fail	Net Lost	Survivor Function	Std. Error	[95% Conf.	Int.]
6	21	3	1	0.8571	0.0764	0.6197	0.9516
7	17	1	0	0.8067	0.0869	0.5631	0.9228
9	16	0	1	0.8067	0.0869	0.5631	0.9228
10	15	1	1	0.7529	0.0963	0.5032	0.8894
11	13	0	1	0.7529	0.0963	0.5032	0.8894
13	12	1	0	0.6902	0.1068	0.4316	0.8491
16	11	1	0	0.6275	0.1141	0.3675	0.8049
17	10	0	1	0.6275	0.1141	0.3675	0.8049
19	9	0	1	0.6275	0.1141	0.3675	0.8049
20	8	0	1	0.6275	0.1141	0.3675	0.8049
22	7	1	0	0.5378	0.1282	0.2678	0.7468
23	6	1	0	0.4482	0.1346	0.1881	0.6801
25	5	0	1	0.4482	0.1346	0.1881	0.6801
32	4	0	2	0.4482	0.1346	0.1881	0.6801
34	2	0	1	0.4482	0.1346	0.1881	0.6801
35	1	0	1	0.4482	0.1346	0.1881	0.6801

SAS Commands for Kaplan Meier Estimator - PROC LIFETEST

The SAS command for the Kaplan-Meier estimate is:

```
        time failtime*censor(1);
or      time failtime*failind(0);
```

The first variable is the failure time, and the second is the failure or censoring indicator. In parentheses you need to put the specific numeric value that corresponds to censoring.

The upper and lower confidence limits on $\hat{S}(t)$ are included in the data set "OUTSURV" when specified. The upper and lower limits are called: **sdf_ucl, sdf_lcl**.

```
data leukemia;
  input weeks remiss;
  label weeks='Time to Remission (in weeks)'
        remiss='Remission indicator (1=yes,0=no)';
  cards;
 6 1
 6 1
.......... ( lines edited out here)
 34 0
 35 0
;

proc lifetest data=leukemia outsurv=confint;
  time weeks*remiss(0);
  title 'Leukemia data from Table 1.1 of Cox and Oakes';
run;

proc print data=confint;
title '95% Confidence Intervals for Estimated Survival';
```

Output from SAS Proc Lifetest

Note: this information is not printed if you use NOPRINT.

Leukemia data from Table 1.1 of Cox and Oakes

The LIFETEST Procedure

Product-Limit Survival Estimates

WEEKS	Survival	Failure	Survival Standard Error	Number Failed	Number Left
0.0000	1.0000	0	0	0	21
6.0000	.	.	.	1	20
6.0000	.	.	.	2	19
6.0000	0.8571	0.1429	0.0764	3	18
6.0000*	.	.	.	3	17
7.0000	0.8067	0.1933	0.0869	4	16
9.0000*	.	.	.	4	15
10.0000	0.7529	0.2471	0.0963	5	14
10.0000*	.	.	.	5	13
11.0000*	.	.	.	5	12
13.0000	0.6902	0.3098	0.1068	6	11
16.0000	0.6275	0.3725	0.1141	7	10
17.0000*	.	.	.	7	9
19.0000*	.	.	.	7	8
20.0000*	.	.	.	7	7
22.0000	0.5378	0.4622	0.1282	8	6
23.0000	0.4482	0.5518	0.1346	9	5
25.0000*	.	.	.	9	4
32.0000*	.	.	.	9	3
32.0000*	.	.	.	9	2
34.0000*	.	.	.	9	1
35.0000*	.	.	.	9	0

* Censored Observation

Output from printing the CONFINT file

95% Confidence Intervals for Estimated Survival

OBS	WEEKS	_CENSOR_	SURVIVAL	SDF_LCL	SDF_UCL
1	0	0	1.00000	1.00000	1.00000
2	6	0	0.85714	0.70748	1.00000
3	6	1	0.85714	.	.
4	7	0	0.80672	0.63633	0.97711
5	9	1	0.80672	.	.
6	10	0	0.75294	0.56410	0.94178
7	10	1	0.75294	.	.
8	11	1	0.75294	.	.
9	13	0	0.69020	0.48084	0.89955
10	16	0	0.62745	0.40391	0.85099
11	17	1	0.62745	.	.
12	19	1	0.62745	.	.
13	20	1	0.62745	.	.
14	22	0	0.53782	0.28648	0.78915
15	23	0	0.44818	0.18439	0.71197
16	25	1	.	.	.
17	32	1	.	.	.
18	32	1	.	.	.
19	34	1	.	.	.
20	35	1	.	.	.

The output dataset will have one observation for each unique combination of WEEKS and _CENSOR_. It will also add an observation for failure time equal to 0.

Splus Commands

Create a file called "leukemia.dat" with the variables names in the first row, as follows:

```
 t    c
 6    1
 6    1
etc ...
```

In Splus, type

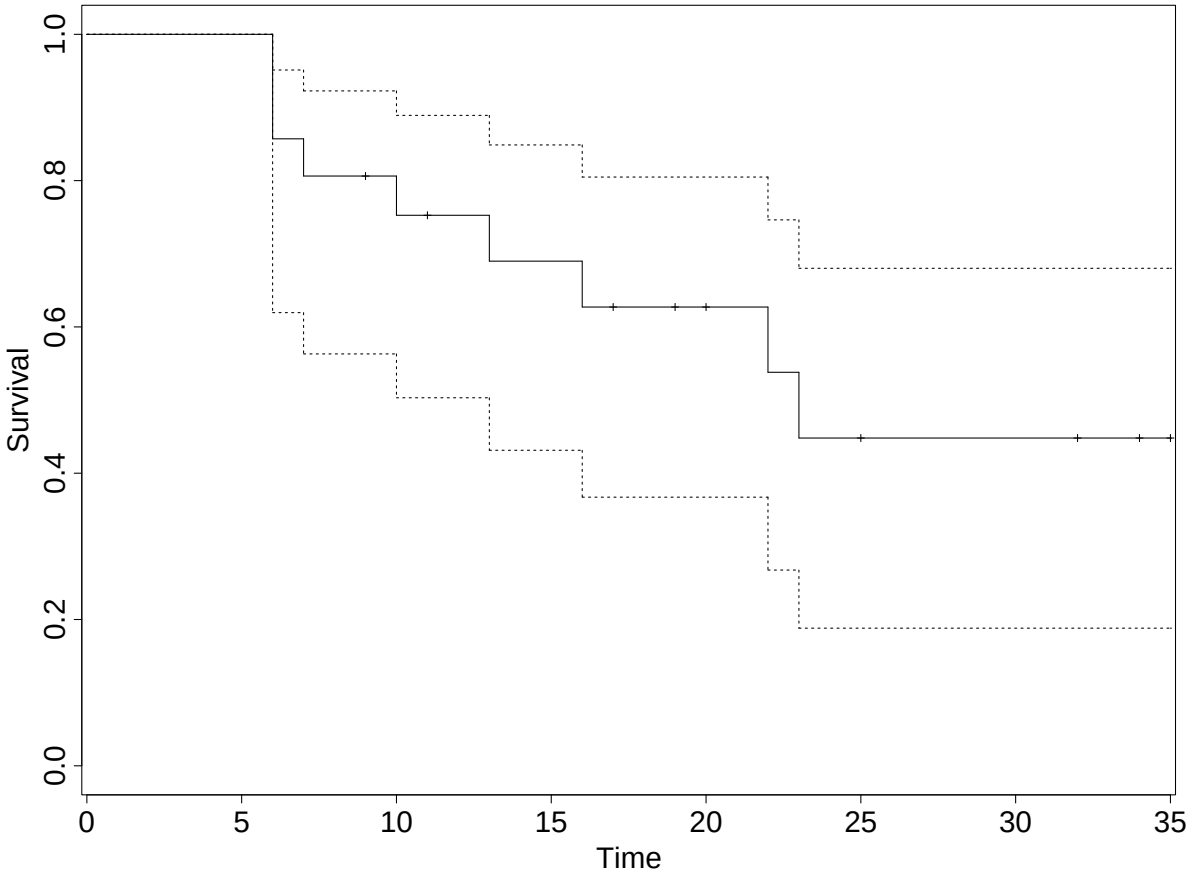
```
y_read.table('leukemia.dat',header=T)
surv.fit(y$t,y$c)
plot(surv.fit(y$t,y$c))
```

(the plot command will also yield 95% confidence intervals)

To specify the type of confidence intervals, use the conf.type= option in the surv.fit statements: e.g. conf.type="log-log" or conf.type="plain"

```
>surv.fit(y$t,y$c)
 95 percent confidence interval is of type "log"
 time n.risk n.event  survival    std.dev lower 95% CI upper 95% CI
    6     21       3 0.8571429 0.07636035    0.7198171    1.0000000
    7     17       1 0.8067227 0.08693529    0.6531242    0.9964437
   10     15       1 0.7529412 0.09634965    0.5859190    0.9675748
   13     12       1 0.6901961 0.10681471    0.5096131    0.9347692
   16     11       1 0.6274510 0.11405387    0.4393939    0.8959949
   22      7       1 0.5378151 0.12823375    0.3370366    0.8582008
   23      6       1 0.4481793 0.13459146    0.2487882    0.8073720


> surv.fit(y$t,y$c,conf.type="log-log")
 95 percent confidence interval is of type "log-log"
 time n.risk n.event  survival    std.dev lower 95% CI upper 95% CI
    6     21       3 0.8571429 0.07636035    0.6197180    0.9515517
    7     17       1 0.8067227 0.08693529    0.5631466    0.9228090
   10     15       1 0.7529412 0.09634965    0.5031995    0.8893618
   13     12       1 0.6901961 0.10681471    0.4316102    0.8490660
   16     11       1 0.6274510 0.11405387    0.3675109    0.8049122
   22      7       1 0.5378151 0.12823375    0.2677789    0.7467907
   23      6       1 0.4481793 0.13459146    0.1880520    0.6801426


> surv.fit(y$t,y$c,conf.type="plain")
 95 percent confidence interval is of type "plain"
 time n.risk n.event  survival    std.dev lower 95% CI upper 95% CI
    6     21       3 0.8571429 0.07636035    0.7074793    1.0000000
    7     17       1 0.8067227 0.08693529    0.6363327    0.9771127
   10     15       1 0.7529412 0.09634965    0.5640993    0.9417830
   13     12       1 0.6901961 0.10681471    0.4808431    0.8995491
   16     11       1 0.6274510 0.11405387    0.4039095    0.8509924
   22      7       1 0.5378151 0.12823375    0.2864816    0.7891487
   23      6       1 0.4481793 0.13459146    0.1843849    0.7119737
```

KM Survival Estimate and Confidence intervals (SPlus)



Means, Medians, Quantiles based on the KM

- **Mean:** $\Sigma_{j=1}^{k} \tau_j \, Pr(T = \tau_j)$

- **Median** - by definition, this is the time, τ, such that $S(\tau) = 0.5$. However, in practice, it is defined as the smallest time such that $\hat{S}(\tau) \leq 0.5$. The median is more appropriate for censored survival data than the mean.

  For the treated leukemia patients, we find:

$$\begin{aligned} \hat{S}(22) &= 0.5378 \\ \hat{S}(23) &= 0.4482 \end{aligned}$$

  The median is thus 23. This can also be seen visually on the graph to the left.

- **Lower quartile (25^{th} percentile):**
  the smallest time (LQ) such that $\hat{S}(LQ) \leq 0.75$

- **Upper quartile (75^{th} percentile):**
  the smallest time (UQ) such that $\hat{S}(UQ) \leq 0.25$

The (2) Lifetable Estimator of Survival:

We said that we would consider the following three methods for estimating a survivorship function

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

without resorting to parametric methods:

(1) $\surd$ **Kaplan-Meier**

(2) $\Longrightarrow$ **Life-table** (Actuarial Estimator)

(3) $\Longrightarrow$ **Cumulative hazard estimator**

(2) The Lifetable or Actuarial Estimator

- one of the oldest techniques around
- used by actuaries, demographers, etc.
- **applies when the data are grouped**

Our goal is still to estimate the survival function, hazard, and density function, but this is complicated by the fact that we don't know exactly when during each time interval an event occurs.

Lee (section 4.2) provides a good description of lifetable methods, and distinguishes several types according to the data sources:

Population Life Tables

- **cohort life table** - describes the mortality experience from birth to death for a particular cohort of people born at about the same time. People at risk at the start of the interval are those who survived the previous interval.
- **current life table** - constructed from (1) census information on the number of individuals alive at each age, for a given year and (2) vital statistics on the number of deaths or failures in a given year, by age. This type of lifetable is often reported in terms of a hypothetical cohort of 100,000 people.

Generally, censoring is not an issue for Population Life Tables.

Clinical Life tables - applies to grouped survival data from studies in patients with specific diseases. Because patients can enter the study at different times, or be lost to follow-up, censoring must be allowed.

Notation

- the j-th time interval is $[t_{j-1}, t_j)$
- c_j - the number of censorings in the j-th interval
- d_j - the number of failures in the j-th interval
- r_j is the number <u>entering</u> the interval

Example: 2418 Males with Angina Pectoris (Lee, p.91)

Year after Diagnosis	j	d_j	c_j	r_j	$r'_j = r_j - c_j/2$	
$[0, 1)$	1	456	0	2418	2418.0	
$[1, 2)$	2	226	39	1962	1942.5	(1962 - $\frac{39}{2}$)
$[2, 3)$	3	152	22	1697	1686.0	
$[3, 4)$	4	171	23	1523	1511.5	
$[4, 5)$	5	135	24	1329	1317.0	
$[5, 6)$	6	125	107	1170	1116.5	
$[6, 7)$	7	83	133	938	871.5	
etc..						

Estimating the survivorship function

We could apply the K-M formula directly to the numbers in the table on the previous page, estimating S(t) as

$$\hat{S}(t) \;=\; \prod_{j:\tau_j<t} \left(1 - \frac{d_j}{r_j}\right)$$

However, this approach is unsatisfactory for grouped data.... it treats the problem as though it were in discrete time, with events happening only at 1 yr, 2 yr, etc. In fact, what we are trying to calculate here is the conditional probability of dying within the interval, given survival to the beginning of it.

What should we do with the censored people?

We can assume that censorings occur:

- at the beginning of each interval: $r'_j = r_j - c_j$
- at the end of each interval: $r'_j = r_j$
- **on average halfway through** the interval:

$$r'_j = r_j - c_j/2$$

The last assumption yields the Actuarial Estimator. It is appropriate if censorings occur uniformly throughout the interval.

Constructing the lifetable

First, some additional notation for the j-th interval, $[t_{j-1}, t_j)$:

- **Midpoint** (t_{mj}) - useful for plotting the density and the hazard function
- **Width** ($b_j = t_j - t_{j-1}$) needed for calculating the hazard in the j-th interval

Quantities estimated:

- Conditional probability of dying

$$\hat{q}_j = d_j / r'_j$$

- Conditional probability of surviving

$$\hat{p}_j = 1 - \hat{q}_j$$

- Cumulative probability of surviving at t_j:

$$\begin{aligned} \hat{S}(t_j) &= \prod_{\ell \leq j} \hat{p}_\ell \\ &= \prod_{\ell \leq j} \left(1 - \frac{d_\ell}{r_{\ell'}}\right) \end{aligned}$$

Some important points to note:

- Because the intervals are defined as $[t_{j-1}, t_j)$, the first interval typically starts with $t_0 = 0$.

- Stata estimates the survival function at the right-hand endpoint of each interval, i.e., $S(t_j)$

- However, SAS estimates the survival function at the left-hand endpoint, $S(t_{j-1})$.

- The implication in SAS is that $\hat{S}(t_0) = 1$ and $\hat{S}(t_1) = p_1$

Other quantities estimated at the midpoint of the j-th interval:

- **Hazard** in the j-th interval:

$$\begin{aligned}\hat{\lambda}(t_{mj}) &= \frac{d_j}{b_j(r'_j - d_j/2)} \\ &= \frac{\hat{q}_j}{b_j(1 - \hat{q}_j/2)}\end{aligned}$$

  the number of deaths in the interval divided by the average number of survivors at the midpoint

- **density** at the midpoint of the j-th interval:

$$\begin{aligned}\hat{f}(t_{mj}) &= \frac{\hat{S}(t_{j-1}) - \hat{S}(t_j)}{b_j} \\ &= \frac{\hat{S}(t_{j-1})\ \hat{q}_j}{b_j}\end{aligned}$$

  Note: Another way to get this is:

$$\begin{aligned}\hat{f}(t_{mj}) &= \hat{\lambda}(t_{mj})\hat{S}(t_{mj}) \\ &= \hat{\lambda}(t_{mj})[\hat{S}(t_j) + \hat{S}(t_{j-1})]/2\end{aligned}$$

Constructing the Lifetable using Stata

Uses the `ltable` command.

If the raw data are already grouped, then the `freq` statement must be used when reading the data.

```
. infile years status count using angina.dat
(32 observations read)

. ltable years status [freq=count]
```

Interval		Beg. Total	Deaths	Lost	Survival	Std. Error	[95% Conf.	Int.]
0	1	2418	456	0	0.8114	0.0080	0.7952	0.8264
1	2	1962	226	39	0.7170	0.0092	0.6986	0.7346
2	3	1697	152	22	0.6524	0.0097	0.6329	0.6711
3	4	1523	171	23	0.5786	0.0101	0.5584	0.5981
4	5	1329	135	24	0.5193	0.0103	0.4989	0.5392
5	6	1170	125	107	0.4611	0.0104	0.4407	0.4813
6	7	938	83	133	0.4172	0.0105	0.3967	0.4376
7	8	722	74	102	0.3712	0.0106	0.3505	0.3919
8	9	546	51	68	0.3342	0.0107	0.3133	0.3553
9	10	427	42	64	0.2987	0.0109	0.2775	0.3201
10	11	321	43	45	0.2557	0.0111	0.2341	0.2777
11	12	233	34	53	0.2136	0.0114	0.1917	0.2363
12	13	146	18	33	0.1839	0.0118	0.1614	0.2075
13	14	95	9	27	0.1636	0.0123	0.1404	0.1884
14	15	59	6	23	0.1429	0.0133	0.1180	0.1701
15	16	30	0	30	0.1429	0.0133	0.1180	0.1701

It is also possible to get estimates of the hazard function, $\hat{\lambda}_j$, and its standard error using the "`hazard`" option:

```
. ltable years status [freq=count], hazard
```

Interval		Beg. Total	Cum. Failure	Std. Error	Hazard	Std. Error	[95% Conf	Int]
0	1	2418	0.1886	0.0080	0.2082	0.0097	0.1892	0.2272
1	2	1962	0.2830	0.0092	0.1235	0.0082	0.1075	0.1396
2	3	1697	0.3476	0.0097	0.0944	0.0076	0.0794	0.1094
3	4	1523	0.4214	0.0101	0.1199	0.0092	0.1020	0.1379
4	5	1329	0.4807	0.0103	0.1080	0.0093	0.0898	0.1262
5	6	1170	0.5389	0.0104	0.1186	0.0106	0.0978	0.1393
6	7	938	0.5828	0.0105	0.1000	0.0110	0.0785	0.1215
7	8	722	0.6288	0.0106	0.1167	0.0135	0.0902	0.1433
8	9	546	0.6658	0.0107	0.1048	0.0147	0.0761	0.1336
9	10	427	0.7013	0.0109	0.1123	0.0173	0.0784	0.1462
10	11	321	0.7443	0.0111	0.1552	0.0236	0.1090	0.2015
11	12	233	0.7864	0.0114	0.1794	0.0306	0.1194	0.2395
12	13	146	0.8161	0.0118	0.1494	0.0351	0.0806	0.2182
13	14	95	0.8364	0.0123	0.1169	0.0389	0.0407	0.1931
14	15	59	0.8571	0.0133	0.1348	0.0549	0.0272	0.2425
15	16	30	0.8571	0.0133	0.0000	.	.	.

There is also a "`failure`" option which gives the number of failures (like the default), and also provides a 95% confidence interval on the cumulative failure probability.

Constructing the lifetable using SAS

If the raw data are already grouped, then the FREQ statement must be used when reading the data.

SAS requires that the interval endpoints be specified, using one of the following (see SAS manual or online help for more detail):

- **intervals** - specify the the interval endpoints
- **width** - specify the width of each interval
- **ninterval** - specify the number of intervals

```
Title 'Actuarial Estimator for Angina Pectoris Example';
data angina;
    input years status count;
cards;
 0.5  1  456
 1.5  1  226
 2.5  1  152                    /* angina cases */
 3.5  1  171
 4.5  1  135
 5.5  1  125
.
.
 0.5  0    0
 1.5  0   39
 2.5  0   22                    /* censored */
 3.5  0   23
 4.5  0   24
 5.5  0  107
.
.

proc lifetest data=angina outsurv=survres intervals=0 to 15 by 1 method=act;
  time years*status(0);
  freq count;
```

SAS output:

Actuarial Estimator for Angina Pectoris Example

The LIFETEST Procedure

Life Table Survival Estimates

Interval [Lower,	Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure	Conditional Probability Standard Error
0	1	456	0	2418.0	0.1886	0.00796
1	2	226	39	1942.5	0.1163	0.00728
2	3	152	22	1686.0	0.0902	0.00698
3	4	171	23	1511.5	0.1131	0.00815
4	5	135	24	1317.0	0.1025	0.00836
5	6	125	107	1116.5	0.1120	0.00944
6	7	83	133	871.5	0.0952	0.00994
7	8	74	102	671.0	0.1103	0.0121
8	9	51	68	512.0	0.0996	0.0132
9	10	42	64	395.0	0.1063	0.0155
10	11	43	45	298.5	0.1441	0.0203
11	12	34	53	206.5	0.1646	0.0258
12	13	18	33	129.5	0.1390	0.0304
13	14	9	27	81.5	0.1104	0.0347
14	15	6	23	47.5	0.1263	0.0482
15	.	0	30	15.0	0	0

Interval [Lower,	Upper)	Survival	Failure	Survival Standard Error	Median Residual Lifetime	Median Standard Error
0	1	1.0000	0	0	5.3313	0.1749
1	2	0.8114	0.1886	0.00796	6.2499	0.2001
2	3	0.7170	0.2830	0.00918	6.3432	0.2361
3	4	0.6524	0.3476	0.00973	6.2262	0.2361
4	5	0.5786	0.4214	0.0101	6.2185	0.1853
5	6	0.5193	0.4807	0.0103	5.9077	0.1806
6	7	0.4611	0.5389	0.0104	5.5962	0.1855
7	8	0.4172	0.5828	0.0105	5.1671	0.2713
8	9	0.3712	0.6288	0.0106	4.9421	0.2763
9	10	0.3342	0.6658	0.0107	4.8258	0.4141
10	11	0.2987	0.7013	0.0109	4.6888	0.4183
11	12	0.2557	0.7443	0.0111	.	.
12	13	0.2136	0.7864	0.0114	.	.
13	14	0.1839	0.8161	0.0118	.	.
14	15	0.1636	0.8364	0.0123	.	.
15	.	0.1429	0.8571	0.0133	.	.

more SAS output: (estimated density $\hat{f}_j$ and hazard $\hat{\lambda}_j$)

Evaluated at the Midpoint of the Interval

Interval [Lower,	Upper)	PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	1	0.1886	0.00796	0.208219	0.009698
1	2	0.0944	0.00598	0.123531	0.008201
2	3	0.0646	0.00507	0.09441	0.007649
3	4	0.0738	0.00543	0.119916	0.009154
4	5	0.0593	0.00495	0.108043	0.009285
5	6	0.0581	0.00503	0.118596	0.010589
6	7	0.0439	0.00469	0.1	0.010963
7	8	0.0460	0.00518	0.116719	0.013545
8	9	0.0370	0.00502	0.10483	0.014659
9	10	0.0355	0.00531	0.112299	0.017301
10	11	0.0430	0.00627	0.155235	0.023602
11	12	0.0421	0.00685	0.17942	0.030646
12	13	0.0297	0.00668	0.149378	0.03511
13	14	0.0203	0.00651	0.116883	0.038894
14	15	0.0207	0.00804	0.134831	0.054919
15	.	.	.	.	.

Summary of the Number of Censored and Uncensored Values

Total	Failed	Censored	%Censored
2418	1625	793	32.7957

Suppose we wish to use the actuarial method, but the data do not come grouped.

Consider the treated nursing home patients, with length of stay (los) grouped into 100 day intervals:

```
.use nurshome

.drop if rx==0                    (keep only the treated patients)
 (881 observations deleted)

.stset los fail

.ltable los fail, intervals(100)
```

Interval		Beg. Total	Deaths	Lost	Survival	Std. Error	[95% Conf.	Int.]
0	100	710	328	0	0.5380	0.0187	0.5006	0.5739
100	200	382	86	0	0.4169	0.0185	0.3805	0.4529
200	300	296	65	0	0.3254	0.0176	0.2911	0.3600
300	400	231	38	0	0.2718	0.0167	0.2396	0.3050
400	500	193	32	1	0.2266	0.0157	0.1966	0.2581
500	600	160	13	0	0.2082	0.0152	0.1792	0.2388
600	700	147	13	0	0.1898	0.0147	0.1619	0.2195
700	800	134	10	30	0.1739	0.0143	0.1468	0.2029
800	900	94	4	29	0.1651	0.0143	0.1383	0.1941
900	1000	61	4	30	0.1508	0.0147	0.1233	0.1808
1000	1100	27	0	27	0.1508	0.0147	0.1233	0.1808

SAS Commands for lifetable analysis - grouping data

```
Title 'Actuarial Estimator for nursing home data';
data morris ;
  infile 'ch12.dat' ;
  input los age trt gender marstat hltstat cens ;

data morristr;
  set morris;
  if trt=1;

proc lifetest data=morristr outsurv=survres
              intervals=0 to 1100 by 100 method=act;
  time los*cens(1);
run ;

proc print data=survres;
run;
```

Actuarial estimator for treated nursing home patients

Actuarial Estimator for Nursing Home Patients

The LIFETEST Procedure

Life Table Survival Estimates

Interval [Lower,	Upper)	Number Failed	Number Censored	Effective Sample Size	Conditional Probability of Failure
0	100	330	0	712.0	0.4635
100	200	86	0	382.0	0.2251
200	300	65	0	296.0	0.2196
300	400	38	0	231.0	0.1645
400	500	32	1	192.5	0.1662
500	600	13	0	160.0	0.0813
600	700	13	0	147.0	0.0884
700	800	10	30	119.0	0.0840
800	900	4	29	79.5	0.0503
900	1000	4	30	46.0	0.0870
1000	1100	0	27	13.5	0

Interval [Lower,	Upper)	Conditional Probability Standard Error	Survival	Failure	Survival Standard Error	Median Residual Lifetime
0	100	0.0187	1.0000	0	0	130.2
100	200	0.0214	0.5365	0.4635	0.0187	306.2
200	300	0.0241	0.4157	0.5843	0.0185	398.8
300	400	0.0244	0.3244	0.6756	0.0175	617.0
400	500	0.0268	0.2711	0.7289	0.0167	.
500	600	0.0216	0.2260	0.7740	0.0157	.
600	700	0.0234	0.2076	0.7924	0.0152	.
700	800	0.0254	0.1893	0.8107	0.0147	.
800	900	0.0245	0.1734	0.8266	0.0143	.
900	1000	0.0415	0.1647	0.8353	0.0142	.
1000	1100	0	0.1503	0.8497	0.0147	.

Actuarial estimator for treated nursing home patients, cont'd

Interval [Lower,	Upper)	Median Standard Error	Evaluated at the Midpoint of the Interval PDF	PDF Standard Error	Hazard	Hazard Standard Error
0	100	15.5136	0.00463	0.000187	0.006033	0.000317
100	200	30.4597	0.00121	0.000122	0.002537	0.000271
200	300	65.7947	0.000913	0.000108	0.002467	0.000304
300	400	74.5466	0.000534	0.000084	0.001792	0.00029
400	500	.	0.000451	0.000078	0.001813	0.000319
500	600	.	0.000184	0.00005	0.000847	0.000235
600	700	.	0.000184	0.00005	0.000925	0.000256
700	800	.	0.000159	0.00005	0.000877	0.000277
800	900	.	0.000087	0.000043	0.000516	0.000258
900	1000	.	0.000143	0.00007	0.000909	0.000454
1000	1100	.	0	.	0	.

Summary of the Number of Censored and Uncensored Values

Total	Failed	Censored	%Censored
712	595	117	16.4326

Actuarial estimator for treated nursing home patients, cont'd
Output from SURVRES dataset

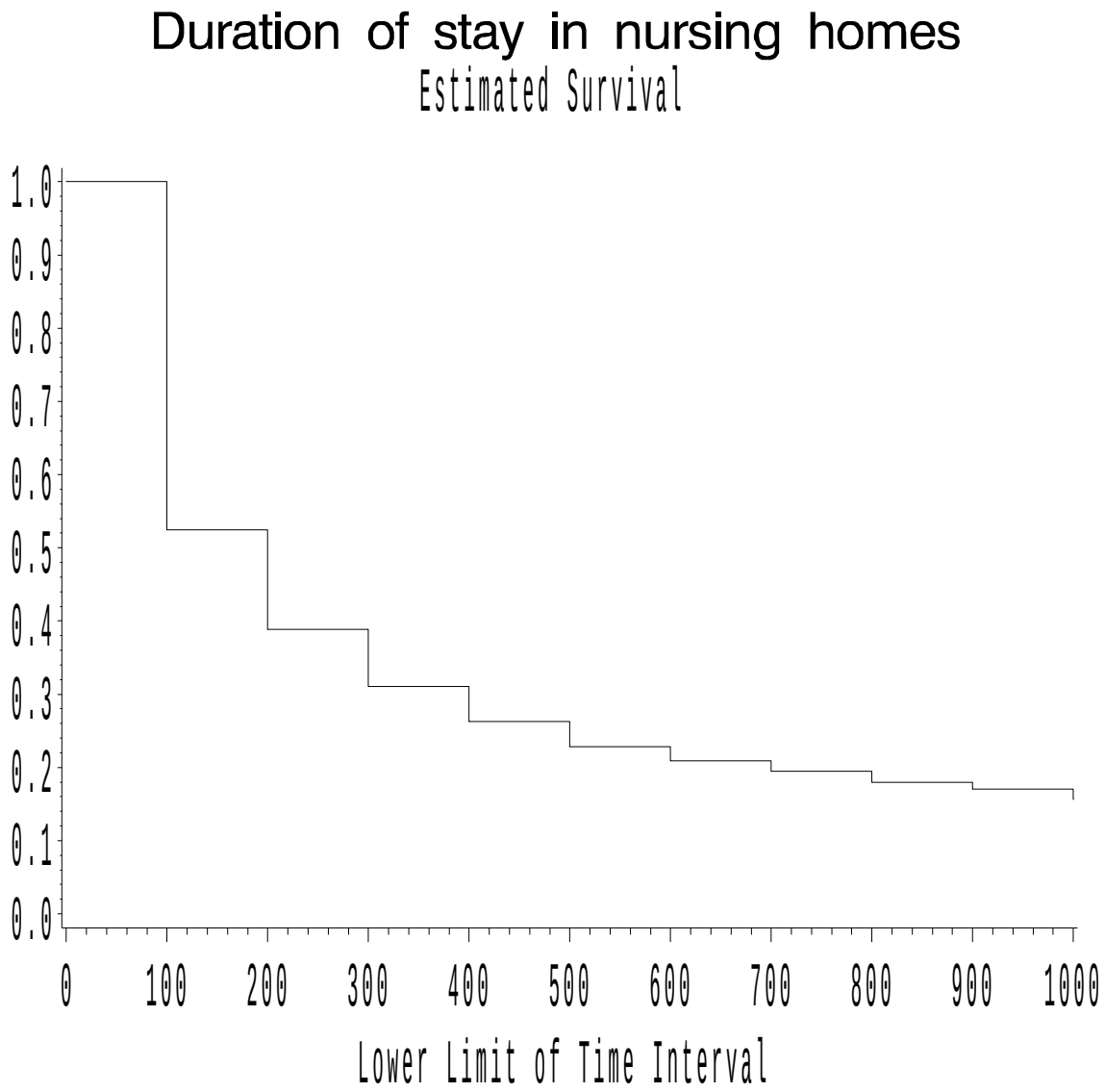
Actuarial Estimator for Nursing Home Patients

OBS	LOS	SURVIVAL	SDF_LCL	SDF_UCL	MIDPOINT	PDF
1	0	1.00000	1.00000	1.00000	50	.0046348
2	100	0.53652	0.49989	0.57315	150	.0012079
3	200	0.41573	0.37953	0.45193	250	.0009129
4	300	0.32444	0.29005	0.35883	350	.0005337
5	400	0.27107	0.23842	0.30372	450	.0004506
6	500	0.22601	0.19528	0.25674	550	.0001836
7	600	0.20764	0.17783	0.23745	650	.0001836
8	700	0.18928	0.16048	0.21808	750	.0001591
9	800	0.17337	0.14536	0.20139	850	.0000872
10	900	0.16465	0.13677	0.19253	950	.0001432
11	1000	0.15033	0.12157	0.17910	1050	.0000000

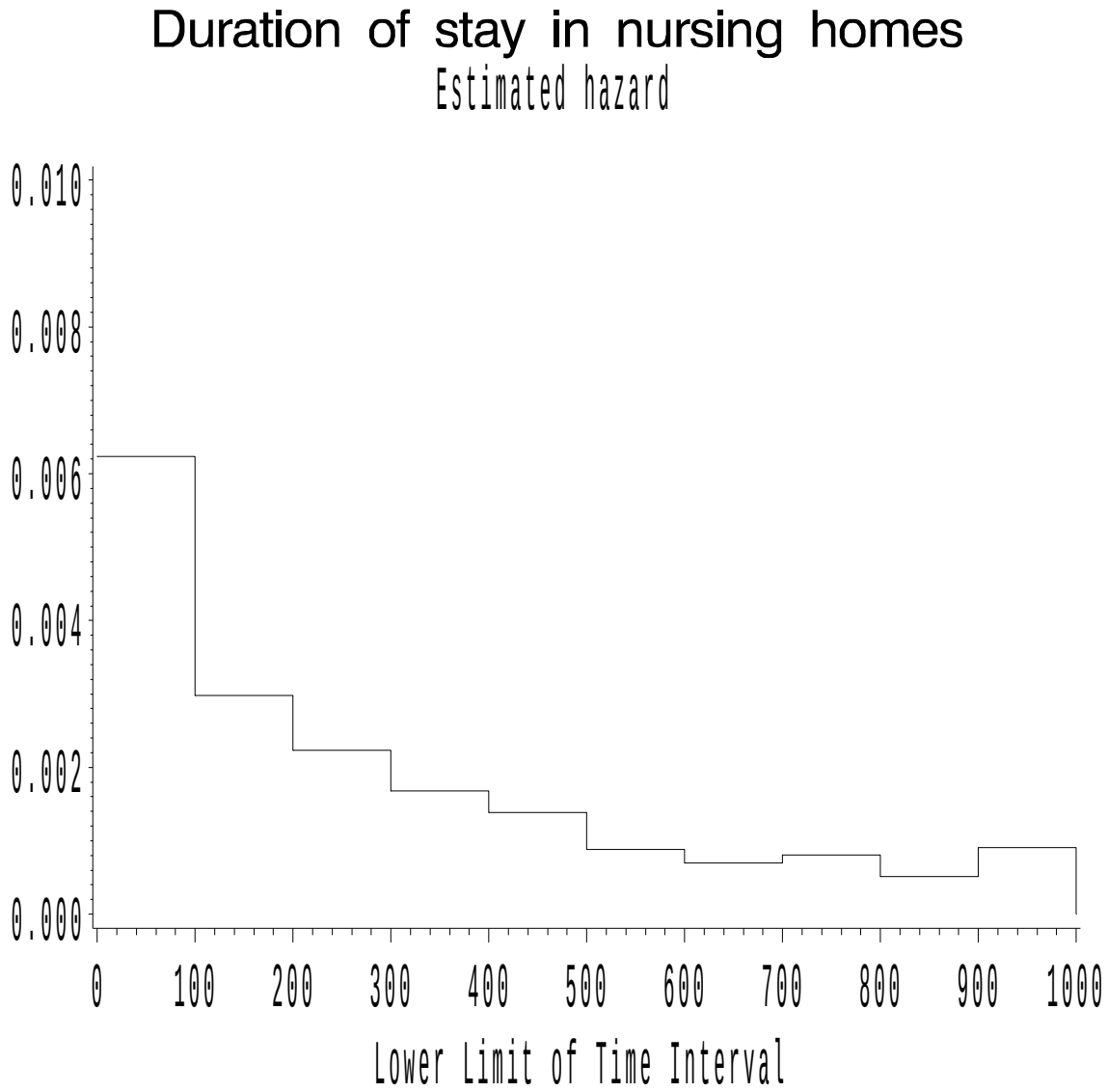
OBS	PDF_LCL	PDF_UCL	HAZARD	HAZ_LCL	HAZ_UCL
1	.0042685	.0050011	.0060329	.0054123	.0066535
2	.0009685	.0014472	.0025369	.0020050	.0030687
3	.0007014	.0011245	.0024668	.0018717	.0030619
4	.0003686	.0006988	.0017925	.0012248	.0023601
5	.0002981	.0006031	.0018130	.0011874	.0024386
6	.0000847	.0002825	.0008469	.0003869	.0013069
7	.0000847	.0002825	.0009253	.0004228	.0014277
8	.0000617	.0002565	.0008772	.0003340	.0014203
9	.0000027	.0001717	.0005161	.0000105	.0010218
10	.0000069	.0002794	.0009091	.0000191	.0017991
11	.	.	.0000000	.	.

Examples for Nursing home data:

Estimated Survival:



Estimated hazard:



(3) Estimating the cumulative hazard

(Nelson-Aalen estimator)

Suppose we want to estimate $\Lambda(t) = \int_0^t \lambda(u)du$, the cumulative hazard at time t.

Just as we did for the KM, think of dividing the observed timespan of the study into a series of fine intervals so that there is only one event per interval:

			D		C		C	D	D	D

$\Lambda(t)$ can then be approximated by a sum:

$$\hat{\Lambda}(t) = \sum_j \lambda_j \Delta$$

where the sum is over intervals, λ_j is the value of the hazard in the j-th interval and Δ is the width of each interval. Since $\hat{\lambda}\Delta$ is approximately the probability of dying in the interval, we can further approximate by

$$\hat{\Lambda}(t) = \sum_j d_j/r_j$$

It follows that $\Lambda(t)$ will change only at death times, and hence we write the Nelson-Aalen estimator as:

$$\hat{\Lambda}_{NA}(t) = \sum_{j:\tau_j<t} d_j/r_j$$

			D		C		C	D	D	D
r_j	n	n	n	n-1	n-1	n-2	n-2	n-3	n-4	
d_j	0	0	1	0	0	0	0	1	1	
c_j	0	0	0	0	1	0	1	0	0	
$\hat{\lambda}(t_j)$	0	0	1/n	0	0	0	0	$\frac{1}{n-3}$	$\frac{1}{n-4}$	
$\hat{\Lambda}(t_j)$	0	0	1/n	1/n	1/n	1/n	1/n			

Once we have $\hat{\Lambda}_{NA}(t)$, we can also find another estimator of $S(t)$ (Fleming-Harrington):

$$\hat{S}_{FH}(t) = \exp(-\hat{\Lambda}_{NA}(t))$$

In general, this estimator of the survival function will be close to the Kaplan-Meier estimator, $\hat{S}_{KM}(t)$

We can also go the other way ... we can take the Kaplan-Meier estimate of $S(t)$, and use it to calculate an alternative estimate of the cumulative hazard function:

$$\hat{\Lambda}_{KM}(t) \;=\; -\log \hat{S}_{KM}(t)$$

Stata commands for FH Survival Estimate

Say we want to obtain the Fleming-Harrington estimate of the survival function for married females, in the healthiest initial subgroup, who are randomized to the untreated group of the nursing home study.

First, we use the following commands to calculate the Nelson-Aalen cumulative hazard estimator:

```
. use nurshome

. keep if rx==0 & gender==0 & health==2 & married==1
 (1579 observations deleted)

. sts list, na

         failure _d:  fail
   analysis time _t:  los

           Beg.          Net   Nelson-Aalen    Std.
  Time    Total   Fail   Lost    Cum. Haz.    Error   [95% Conf. Int.]
-------------------------------------------------------------------------
    14       12      1      0       0.0833   0.0833   0.0117    0.5916
    24       11      1      0       0.1742   0.1233   0.0435    0.6976
    25       10      1      0       0.2742   0.1588   0.0882    0.8530
    38        9      1      0       0.3854   0.1938   0.1438    1.0326
    64        8      1      0       0.5104   0.2306   0.2105    1.2374
    89        7      1      0       0.6532   0.2713   0.2894    1.4742
   113        6      1      0       0.8199   0.3184   0.3830    1.7551
   123        5      1      0       1.0199   0.3760   0.4952    2.1006
   149        4      1      0       1.2699   0.4515   0.6326    2.5493
   168        3      1      0       1.6032   0.5612   0.8073    3.1840
   185        2      1      0       2.1032   0.7516   1.0439    4.2373
   234        1      1      0       3.1032   1.2510   1.4082    6.8384
-------------------------------------------------------------------------
```

After generating the Nelson-Aalen estimator, we manually have to create a variable for the survival estimate:

```
. sts gen nelson=na

. gen sfh=exp(-nelson)

. list sfh

          sfh
  1.  .9200444
  2.  .8400932
  3.  .7601478
  4.  .6802101
  5.  .6002833
  6.  .5203723
  7.  .4404857
  8.  .3606392
  9.  .2808661
 10.  .2012493
 11.  .1220639
 12.  .0449048
```

Additional built-in functions can be used to generate 95% confidence intervals on the FH survival estimate.

We can compare the Fleming-Harrington survival estimate to the KM estimate by rerunning the `sts list` command:

```
. sts list

. sts gen skm=s

. list skm sfh

            skm        sfh
  1.   .91666667   .9200444
  2.   .83333333   .8400932
  3.         .75   .7601478
  4.   .66666667   .6802101
  5.   .58333333   .6002833
  6.          .5   .5203723
  7.   .41666667   .4404857
  8.   .33333333   .3606392
  9.         .25   .2808661
 10.   .16666667   .2012493
 11.   .08333333   .1220639
 12.           0   .0449048
```

In this example, it looks like the Fleming-Harrington estimator is slightly higher than the KM at every time point, but with larger datasets the two will typically be much closer.

Splus Commands for Fleming-Harrington Estimator:

(Nursing home data: females, untreated, married, healthy)

Fleming-Harrington:

```
>fh<-surv.fit(los,cens,type="f",conf.type="log-log")
>fh

 95 percent confidence interval is of type "log-log"
 time n.risk n.event   survival    std.dev lower 95% CI upper 95% CI
   14     12       1 0.9200444 0.08007959 0.5244209125    0.9892988
   24     11       1 0.8400932 0.10845557 0.4750041174    0.9600371
   25     10       1 0.7601478 0.12669130 0.4055610500    0.9200425
   38      9       1 0.6802101 0.13884731 0.3367907188    0.8724502
   64      8       1 0.6002833 0.14645413 0.2718422278    0.8187596
   89      7       1 0.5203723 0.15021856 0.2115701242    0.7597900
  113      6       1 0.4404857 0.15045450 0.1564397006    0.6960354
  123      5       1 0.3606392 0.14723033 0.1069925657    0.6278888
  149      4       1 0.2808661 0.14043303 0.0640979523    0.5560134
  168      3       1 0.2012493 0.12990589 0.0293208029    0.4827590
  185      2       1 0.1220639 0.11686728 0.0058990525    0.4224087
  234      1       1 0.0449048 0.06216787 0.0005874321    0.2740658
```

Kaplan-Meier:

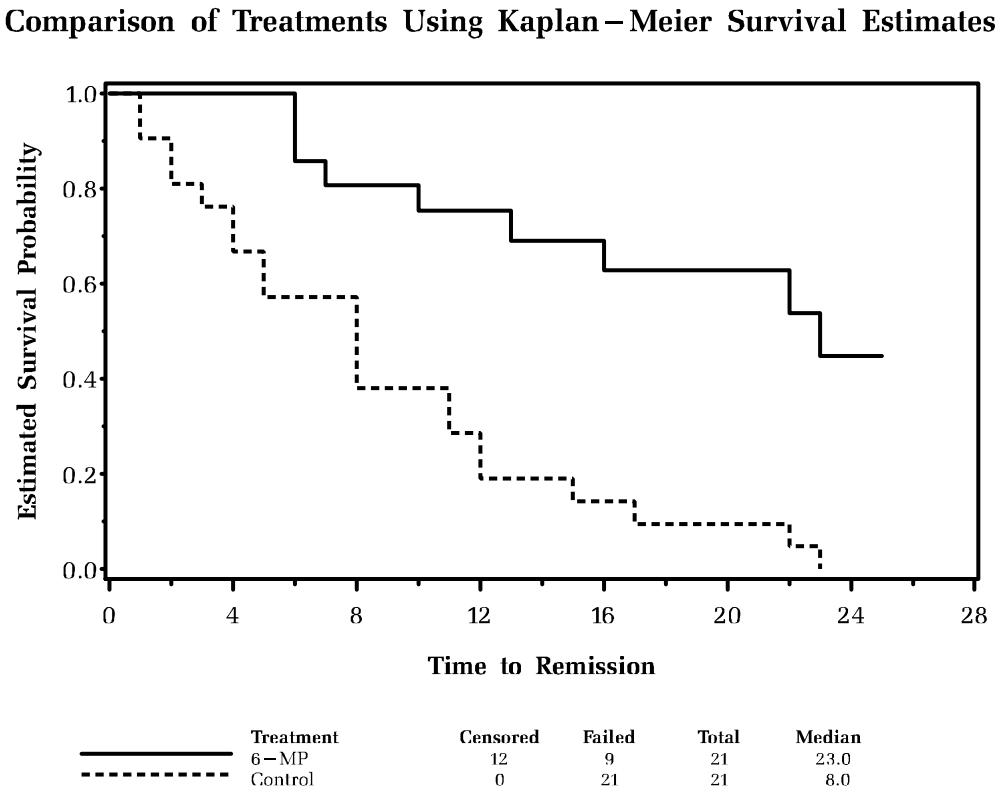
```
>km<-surv.fit(los,cens,conf.type="log-log")
>km

 95 percent confidence interval is of type "log-log"
 time n.risk n.event  survival    std.dev lower 95% CI upper 95% CI
   14     12       1 0.91666667 0.07978559  0.538977181    0.9878256
   24     11       1 0.83333333 0.10758287  0.481714942    0.9555094
   25     10       1 0.75000000 0.12500000  0.408415913    0.9117204
   38      9       1 0.66666667 0.13608276  0.337018933    0.8597118
   64      8       1 0.58333333 0.14231876  0.270138924    0.8009402
   89      7       1 0.50000000 0.14433757  0.208477143    0.7360731
  113      6       1 0.41666667 0.14231876  0.152471264    0.6653015
  123      5       1 0.33333333 0.13608276  0.102703980    0.5884189
  149      4       1 0.25000000 0.12500000  0.060144556    0.5047588
  168      3       1 0.16666667 0.10758287  0.026510427    0.4129803
  185      2       1 0.08333333 0.07978559  0.005052835    0.3110704
  234      1       1 0.00000000          NA           NA           NA
```

Comparison of Survival Curves

We spent the last class looking at some nonparametric approaches for estimating the survival function, $\hat{S}(t)$, over time for a single sample of individuals.

Now we want to compare the survival estimates between two groups.

Example: Time to remission of leukemia patients



How can we form a basis for comparison?

At a specific point in time, we could see whether the confidence intervals for the survival curves overlap.

However, the confidence intervals we have been calculating are **"pointwise"** $\Rightarrow$ they correspond to a confidence interval for $\hat{S}(t^*)$ at a single point in time, t^*.

In other words, we can't say that the true survival function $S(t)$ is contained between the pointwise confidence intervals with 95% probability.

(**Aside:** if you're interested, the issue of confidence **bands** for the estimated survival function are discussed in Section 4.4 of Klein and Moeschberger)

Looking at whether the confidence intervals for $\hat{S}(t^*)$ overlap between the 6MP and placebo groups would only focus on comparing the two treatment groups at a single point in time, t^*. We want an overall comparison.

Should we base our overall comparison of $\hat{S}(t)$ on:

- the furthest distance between the two curves?
- the median survival for each group?
- the average hazard? (for exponential distributions, this would be like comparing the mean event times)
- adding up the difference between the two survival estimates over time?

$$\sum_j \left[\hat{S}(t_{jA}) - \hat{S}(t_{jB})\right]$$

- a weighted sum of differences, where the weights reflect the number at risk at each time?
- a rank-based test? i.e., we could rank all of the event times, and then see whether the sum of ranks for one group was less than the other.

Nonparametric comparisons of groups

All of these are pretty reasonable options, and we'll see that there have been several proposals for how to compare the survival of two groups. For the moment, we are sticking to nonparametric comparisons.

Why nonparametric?

- **fairly robust**
- **efficient relative to parametric tests**
- **often simple and intuitive**

Before continuing the description of the two-sample comparison, I'm going to try to put this in a general framework to give a perspective of where we're heading in this class.

General Framework for Survival Analysis

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

- X_i is a censored failure time random variable
- δ_i is the failure/censoring indicator
- $\mathbf{Z}_i$ represents a set of covariates

Note that $\mathbf{Z}_i$ might be a scalar (a single covariate, say treatment or gender) or may be a $(p \times 1)$ vector (representing several different covariates).

These covariates might be:

- continuous
- discrete
- time-varying (more later)

If $\mathbf{Z}_i$ is a scalar and is binary, then we are comparing the survival of two groups, like in the leukemia example.

More generally though, it is useful to build a model that characterizes the relationship between survival and all of the covariates of interest.

We'll proceed as follows:

- Two group comparisons
- Multigroup and stratified comparisons - stratified logrank
- Failure time regression models
  - Cox proportional hazards model
  - Accelerated failure time model

Two sample tests

- Mantel-Haenszel logrank test
- Peto & Peto's version of the logrank test
- Gehan's Generalized Wilcoxon
- Peto & Peto's and Prentice's generalized Wilcoxon
- Tarone-Ware and Fleming-Harrington classes
- Cox's F-test (non-parametric version)

References:

Hosmer & Lemeshow	Section 2.4
Collett	Section 2.5
Klein & Moeschberger	Section 7.3
Kleinbaum	Chapter 2
Lee	Chapter 5

Mantel-Haenszel Logrank test

The logrank test is the most well known and widely used.

It also has an intuitive appeal, building on standard methods for binary data. (Later we will see that it can also be obtained as the score test from a partial likelihood from the Cox Proportional Hazards model.)

First consider the following (2×2) table classifying those with and without the event of interest in a two group setting:

	Event		
Group	Yes	No	Total
0	d_0	$n_0 - d_0$	n_0
1	d_1	$n_1 - d_1$	n_1
Total	d	$n - d$	n

If the margins of this table are considered fixed, then d_0 follows a ?distribution. Under the null hypothesis of no association between the event and group, it follows that

$$E(d_0) = \frac{n_0 d}{n}$$

$$Var(d_0) = \frac{n_0\, n_1\, d(n-d)}{n^2(n-1)}$$

Therefore, under H_0:

$$\chi^2_{MH} = \frac{[d_0 - n_0\, d/n]^2}{\frac{n_0\, n_1\, d(n-d)}{n^2(n-1)}} \sim \chi^2_1$$

This is the Mantel-Haenszel statistic and is approximately equivalent to the Pearson χ^2 test for equality of the two groups given by:

$$\chi^2_p = \sum \frac{(o-e)^2}{e}$$

Note: recall that the Pearson χ^2 test was derived for the case where only the row margins were fixed, and thus the variance above was replaced by:

$$Var(d_0 - \frac{n_0(d_0+d_1)}{n}) = \frac{n_0\, n_1\, d(n-d)}{n^3}$$

Example: Toxicity in a clinical trial with two treatments

	Toxicity		
Group	Yes	No	Total
0	8	42	50
1	2	48	50
Total	10	90	100

$$\chi^2_p = 4.00 \quad (p = 0.046)$$

$$\chi^2_{MH} = 3.96 \quad (p = 0.047)$$

Now suppose we have K (2×2) tables, all independent, and we want to test for a common group effect. The Cochran-Mantel-Haenszel test for a common odds ratio not equal to 1 can be written as:

$$\chi^2_{CMH} = \frac{[\Sigma_{j=1}^{K}(d_{0j} - n_{0j} * d_j/n_j)]^2}{\Sigma_{j=1}^{K} n_{1j}n_{0j}d_j(n_j - d_j)/[n_j^2(n_j - 1)]}$$

where the subscript j refers to the j-th table:

	Event		
Group	Yes	No	Total
0	d_{0j}	$n_{0j} - d_{0j}$	n_{0j}
1	d_{1j}	$n_{1j} - d_{1j}$	n_{1j}
Total	d_j	$n_j - d_j$	n_j

This statistic is distributed approximately as χ^2_1.

How does this apply in survival analysis?

Suppose we observe

Group 1: $(X_{11}, \delta_{11}) \ldots (X_{1n_1}, \delta_{1n_1})$

Group 0: $(X_{01}, \delta_{01}) \ldots (X_{0n_0}, \delta_{0n_0})$

We could just count the numbers of failures: eg., $d_1 = \Sigma_{j=1}^{K} \delta_{1j}$

Example: Leukemia data, just counting up the number of remissions in each treatment group.

	Fail		
Group	Yes	No	Total
0	21	0	21
1	9	12	21
Total	30	12	42

$$\begin{aligned} \chi^2_p &= 16.8 \quad (p = 0.001) \\ \chi^2_{MH} &= 16.4 \quad (p = 0.001) \end{aligned}$$

But, this doesn't account for the time at risk.

Conceptually, we would like to compare the KM survival curves. Let's put the components side-by-side and compare.

Cox & Oakes Table 1.1 Leukemia example

Ordered	Group 0			Group 1		
Death Times	d_j	c_j	r_j	d_j	c_j	r_j
1	2	0	21	0	0	21
2	2	0	19	0	0	21
3	1	0	17	0	0	21
4	2	0	16	0	0	21
5	2	0	14	0	0	21
6	0	0	12	3	1	21
7	0	0	12	1	0	17
8	4	0	12	0	0	16
9	0	0	8	0	1	16
10	0	0	8	1	1	15
11	2	0	8	0	1	13
12	2	0	6	0	0	12
13	0	0	4	1	0	12
15	1	0	4	0	0	11
16	0	0	3	1	0	11
17	1	0	3	0	1	10
19	0	0	2	0	1	9
20	0	0	2	0	1	8
22	1	0	2	1	0	7
23	1	0	1	1	0	6
25	0	0	0	0	1	5

Note that I wrote down the number at risk for Group 1 for times 1-5 even though there were no events or censorings at those times.

Logrank Test: Formal Definition

The logrank test is obtained by constructing a (2×2) table at each distinct death time, and comparing the death rates between the two groups, conditional on the number at risk in the groups. The tables are then combined using the Cochran-Mantel-Haenszel test.

Note: The logrank is sometimes called the Cox-Mantel test.

Let $t_1, ..., t_K$ represent the K ordered, distinct death times. At the j-th death time, we have the following table:

	Die/Fail		
Group	Yes	No	Total
0	d_{0j}	$r_{0j} - d_{0j}$	r_{0j}
1	d_{1j}	$r_{1j} - d_{1j}$	r_{1j}
Total	d_j	$r_j - d_j$	r_j

where d_{0j} and d_{1j} are the number of deaths in group 0 and 1, respectively at the j-th death time, and r_{0j} and r_{1j} are the number at risk at that time, in groups 0 and 1.

The logrank test is:

$$\chi^2_{logrank} = \frac{[\Sigma_{j=1}^{K}(d_{0j} - r_{0j} * d_j/r_j)]^2}{\Sigma_{j=1}^{K} \frac{r_{1j} r_{0j} d_j (r_j - d_j)}{[r_j^2 (r_j - 1)]}}$$

Assuming the tables are all independent, then this statistic will have an approximate χ^2 distribution with 1 df.

Based on the motivation for the logrank test, which of the survival-related quantities are we comparing at each time point?

- $\Sigma_{j=1}^{K} w_j \left[\hat{S}_1(t_j) - \hat{S}_2(t_j)\right]$ **?**
- $\Sigma_{j=1}^{K} w_j \left[\hat{\lambda}_1(t_j) - \hat{\lambda}_2(t_j)\right]$ **?**
- $\Sigma_{j=1}^{K} w_j \left[\hat{\Lambda}_1(t_j) - \hat{\Lambda}_2(t_j)\right]$ **?**

First several tables of leukemia data

CMH analysis of leukemia data

TABLE 1 OF TRTMT BY REMISS
CONTROLLING FOR FAILTIME=1

TRTMT REMISS

Frequency| Expected	0	1	Total
0	19	2	21
	20	1	
1	21	0	21
	20	1	
Total	40	2	42

TABLE 2 OF TRTMT BY REMISS
CONTROLLING FOR FAILTIME=2

TRTMT REMISS

Frequency| Expected	0	1	Total
0	17	2	19
	18.05	0.95	
1	21	0	21
	19.95	1.05	
Total	38	2	40

TABLE 3 OF TRTMT BY REMISS
CONTROLLING FOR FAILTIME=3

TRTMT REMISS

Frequency| Expected	0	1	Total
0	16	1	17
	16.553	0.4474	
1	21	0	21
	20.447	0.5526	
Total	37	1	38

TABLE 4 OF TRTMT BY REMISS
CONTROLLING FOR FAILTIME=4

TRTMT REMISS

Frequency| Expected	0	1	Total
0	14	2	16
	15.135	0.8649	
1	21	0	21
	19.865	1.1351	
Total	35	2	37

CMH statistic = logrank statistic

SUMMARY STATISTICS FOR TRTMT BY REMISS
CONTROLLING FOR FAILTIME

Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic	Alternative Hypothesis	DF	Value	Prob	
1	Nonzero Correlation	1	16.793	0.001	
2	Row Mean Scores Differ	1	16.793	0.001	
3	General Association	1	16.793	0.001	<===LOGRANK TEST

Note: Although CMH works to get the correct logrank test, it would require inputting the d_j and r_j at each time of death for each treatment group. There's an easier way to get the test statistic, which I'll show you shortly.

Calculating logrank statistic by hand
Leukemia Example:

Ordered	Group 0		Combined				
Death Times	d_{0j}	r_{0j}	d_j	r_j	e_j	$o_j - e_j$	v_j
1	2	21	2	42	1.00	1.00	0.488
2	2	19	2	40	0.95	1.05	
3	1	17	1	38	0.45	0.55	
4	2	16	2	37	0.86	1.14	
5	2	14	2	35			
6	0	12	3	33			
7	0	12	1	29			
8	4	12	4	28			
10	0	8	1	23			
11	2	8	2	21			
12	2	6	2	18			
13	0	4	1	16			
15	1	4	1	15			
16	0	3	1	14			
17	1	3	1	13			
22	1	2	2	9			
23	1	1	2	7			
Sum						10.251	6.257

$o_j = d_{0j}$

$e_j = d_j r_{0j}/r_j$

$v_j = r_{1j} r_{0j} d_j (r_j - d_j)/[r_j^2 (r_j - 1)]$

$$\chi^2_{logrank} = \frac{(10.251)^2}{6.257} = 16.793$$

Notes about logrank test:

- The logrank statistic depends on ranks of event times only

- If there are no tied deaths, then the logrank has the form:

$$\frac{[\Sigma_{j=1}^{K}(d_{0j} - \frac{r_{0j}}{r_j})]^2}{\Sigma_{j=1}^{K} r_{1j} r_{0j}/r_j^2}$$

- Numerator can be interpreted as $\Sigma(o - e)$ where "o" is the observed number of deaths in group 0, and "e" is the expected number, given the risk set. The expected number equals #deaths $\times$ proportion in group 0 at risk.

- The $(o - e)$ terms in the numerator can be written as

$$\frac{r_{0j} r_{1j}}{r_j}(\hat{\lambda}_{1j} - \hat{\lambda}_{0j})$$

- It does not matter which group you choose to sum over.

  To see this, note that if we summed up (o-e) over the death times for the 6MP group we would get -10.251, and the sum of the variances is the same. So when we square the numerator, the test statistic is the same.

Analogous to the CMH test for a series of tables at different levels of a confounder, the logrank test is most powerful when "odds ratios" are constant over time intervals. That is, it is most powerful for **proportional hazards**.

Checking the assumption of proportional hazards:

- check to see if the estimated survival curves cross - if they do, then this is evidence that the hazards are <u>not</u> proportional
- more formal test: **any ideas?**

What should be done if the hazards are <u>not</u> proportional?

- If the difference between hazards has a consistent sign, the logrank test usually does well.
- Other tests are available that are more powerful against different alternatives.

Getting the logrank statistic using Stata:

After declaring data as survival type data using the "stset" command, issue the "sts test" command

```
. stset remiss status

   data set name:  leukem
              id:  --               (meaning each record a unique subject)
      entry time:  --               (meaning all entered at time 0)
       exit time:  remiss
  failure/censor:  status

. sts list, by(trt)

           Beg.          Net   Survivor      Std.
  Time    Total   Fail   Lost  Function     Error     [95% Conf. Int.]
-------------------------------------------------------------------------
trt=0
     1       21      2      0    0.9048    0.0641     0.6700    0.9753
     2       19      2      0    0.8095    0.0857     0.5689    0.9239
     3       17      1      0    0.7619    0.0929     0.5194    0.8933
     4       16      2      0    0.6667    0.1029     0.4254    0.8250
.
.   (etc)

. sts test trt

Log-rank test for equality of survivor functions
------------------------------------------------
      |  Events
trt   |  observed       expected
------+-------------------------
0     |        21          10.75
1     |         9          19.25
------+-------------------------
Total |        30          30.00

            chi2(1) =      16.79
            Pr>chi2 =     0.0000
```

Getting the logrank statistic using SAS

- Still use PROC LIFETEST
- Add "STRATA" command, with treatment variable
- Gives the chi-square test (2-sided), but also gives you the terms you need to calculate the 1-sided test; this is useful if we want to know which of the two groups has the higher estimated hazard over time.
- The STRATA command also gives the Gehan-Wilcoxon test (which we will talk about next)

```
Title 'Cox and Oakes example';
data leukemia;
    input weeks remiss trtmt;
    cards;
6     0     1
6     1     1
6     1     1
6     1     1        /* data for 6MP group */
7     1     1
9     0     1
etc
1     1     0
1     1     0        /* data for placebo group */
2     1     0
2     1     0
etc
;

proc lifetest data=leukemia;
  time weeks*remiss(0);
  strata trtmt;
  title 'Logrank test for leukemia data';
run;
```

Output from leukemia example:

```
Logrank test for leukemia data

Summary of the Number of Censored and Uncensored Values

TRTMT       Total     Failed   Censored  %Censored

6-MP           21          9         12    57.1429
Control        21         21          0     0.0000

Total          42         30         12    28.5714


Testing Homogeneity of Survival Curves over Strata
Time Variable FAILTIME

Rank Statistics

TRTMT        Log-Rank     Wilcoxon

6-MP          -10.251      -271.00
Control        10.251       271.00


Covariance Matrix for the Log-Rank Statistics

TRTMT             6-MP       Control

6-MP           6.25696      -6.25696
Control       -6.25696       6.25696


Test of Equality over Strata
                                 Pr >
Test      Chi-Square    DF  Chi-Square

Log-Rank     16.7929     1      0.0001  <== Here's the one we want!!
Wilcoxon     13.4579     1      0.0002
-2Log(LR)    16.4852     1      0.0001
```

Getting the logrank statistic using Splus:

Instead of the "surv.fit" command, use the "surv.diff" command with a "group" (treatment) variable.

```
Mantel-Haenszel logrank:

> logrank<-surv.diff(weeks,remiss,trtmt)
> logrank
      N Observed Expected (O-E)^2/E
   0 21       21    10.75     9.775
   1 21        9    19.25     5.458

    Chisq= 16.8  on 1 degrees of freedom, p= 4.169e-05
```

Generalization of logrank test
$\Longrightarrow$ Linear rank tests

The logrank and other tests can be derived by assigning scores to the ranks of the death times, and are members of a general class of **linear rank tests** (for more detail, see Lee, ch 5)

First, define

$$\hat{\Lambda}(t) = \sum_{j:t_j<t} \frac{d_j}{r_j}$$

where d_j and r_j are the number of deaths and the number at risk, respectively at the j-th ordered death time.

Then assign these scores (suggested by Peto and Peto):

Event	Score
Death at t_j	$w_j = 1 - \hat{\Lambda}(t_j)$
Censoring at t_j	$w_j = -\hat{\Lambda}(t_j)$

To calculate the logrank test, simply sum up the scores for group 0.

Example Group 0: 15, 18, 19, 19, 20

Group 1: 16+, 18+, 20+, 23, 24+

Calculation of logrank as a linear rank statistic

Ordered Data	Group	d_j	r_j	$\hat{\Lambda}(t_j)$	score w_j
15	0	1	10	0.100	0.900
16^+	1	0	9	0.100	-0.100
18	0	1	8	0.225	0.775
18^+	1	0	7	0.225	-0.225
19	0	2	6	0.558	0.442
20	0	1	4	0.808	0.192
20^+	1	0	3	0.808	-0.808
23	1	1	2	1.308	-0.308
24^+	1	0	1	1.308	-1.308

The logrank statistic S is sum of scores for group 0:

$$S = 0.900 + 0.775 + 0.442 + 0.442 + 0.192 = 2.75$$

The variance is:

$$Var(S) = \frac{n_0 n_1 \, \Sigma_{j=1}^{n} \, w_j^2}{n(n-1)}$$

In this case, $Var(S) = 1.210$, so

$$Z = \frac{2.75}{\sqrt{1.210}} = 2.50 \implies \chi^2_{logrank} = (2.50)^2 = 6.25$$

Why is this form of the logrank equivalent?

The logrank statistic S is equivalent to $\Sigma(o - e)$ over the distinct death times, where "o" is the observed number of deaths in group 0, and "e" is the expected number, given the risk sets.

At deaths:	weights are $1 - \hat{\Lambda}$
At censorings:	weights are $-\hat{\Lambda}$

So we are summing up "1's" for deaths (to get d_{0j}), and subtracting $-\hat{\Lambda}$ at both deaths and censorings. This amounts to subtracting d_j/r_j at each death or censoring time in group 0, at or after the j-th death. Since there are a total of r_{0j} of these, we get $e = r_{0j} * d_j/r_j$.

Why is it called the **logrank** test?

Since $S(t) = \exp(-\Lambda(t))$, an alternative estimator of $S(t)$ is:

$$\hat{S}(t) = \exp(-\hat{\Lambda}(t)) = \exp(-\sum_{j:t_j<t} \frac{d_j}{r_j})$$

So, we can think of $\hat{\Lambda}(t) = -\log(\hat{S}(t))$ as yielding the "log-survival" scores used to calculate the statistic.

Comparing the CMH-type Logrank and "Linear Rank" logrank

A. CMH-type Logrank:

We motivated the logrank test through the CMH statistic for testing $H_o : OR = 1$ over K tables, where K is the number of distinct death times. This turned out to be what we get when we use the logrank (default) option in Stata or the "STRATA" statement in SAS.

B. Linear Rank logrank:

The linear rank version of the logrank test is based on adding up "scores" for one of the two treatment groups. The particular scores that gave us the same logrank statistic were based on the Nelson-Aalen estimator, i.e., $\hat{\Lambda} = \Sigma\, \hat{\lambda}(t_j)$. This is what you get when you use the "TEST" statement in SAS.

Here are some comparisons, with a new example to show when the two types of logrank statistics will be equal.

First, let's go back to our example from Chapter 5 of Lee:

Example Group 0: 15, 18, 19, 19, 20

Group 1: 16+, 18+, 20+, 23, 24+

A. The CMH-type logrank statistic: (using the STRATA statement)

```
Rank Statistics

TRTMT        Log-Rank     Wilcoxon

Control        2.7500       18.000
Treated       -2.7500      -18.000


Covariance Matrix for the Log-Rank Statistics

TRTMT          Control       Treated

Control        1.08750      -1.08750
Treated       -1.08750       1.08750


Test of Equality over Strata

                                 Pr >
Test      Chi-Square    DF  Chi-Square

Log-Rank      6.9540     1      0.0084
Wilcoxon      5.5479     1      0.0185
-2Log(LR)     3.3444     1      0.0674
```

This is exactly the same chi-square test that you would get if you calculated the numerator of the logrank as $\Sigma(o_j - e_j)$ and the variance as $v_j = r_{1j} r_{0j} d_j (r_j - d_j)/[r_j^2 (r_j - 1)]$

Ordered	Group 0		Combined				
Death Times	d_{0j}	r_{0j}	d_j	r_j	e_j	$o_j - e_j$	v_j
15	1	5	1	10	0.50	0.50	0.2500
18	1	4	1	8	0.50	0.50	0.2500
19	2	3	2	6	1.00	1.00	0.4000
20	1	1	2	4	0.25	0.75	0.1870
23	0	0	1	2	0.00	0.00	0.0000
Sum						2.75	1.0875

$$\chi^2_{logrank} = \frac{(2.75)^2}{1.0875} = 6.954$$

B. The "linear rank" logrank statistic:
(using the TEST statement)

```
Univariate Chi-Squares for the LOG RANK Test

              Test       Standard                    Pr >
Variable    Statistic    Deviation    Chi-Square    Chi-Square

GROUP         2.7500       1.0897       6.3684       0.0116


Covariance Matrix for the LOG RANK Statistics

Variable         TRTMT

TRTMT          1.18750
```

This is actually very close to what we would get if we use the Nelson-Aalen based "scores":

Calculation of logrank as a linear rank statistic

Ordered Data	Group	d_j	r_j	$\hat{\Lambda}(t_j)$	score w_j
15	0	1	10	0.100	0.900
16^+	1	0	9	0.100	-0.100
18	0	1	8	0.225	0.775
18^+	1	0	7	0.225	-0.225
19	0	2	6	0.558	0.442
20	0	1	4	0.808	0.192
20^+	1	0	3	0.808	-0.808
23	1	1	2	1.308	-0.308
24^+	1	1	1	1.308	-1.308
Sum(grp 0)					2.750

Note that the numerator is the exact same number (2.75) in both versions of the logrank test. The difference in the denominator is due to the way that ties are handled.

CMH-type variance:

$$\begin{aligned} var &= \sum \frac{r_{1j} r_{0j} d_j (r_j - d_j)}{r_j^2 (r_j - 1)} \\ &= \sum \frac{r_{1j} r_{0j}}{r_j (r_j - 1)} \frac{d_j (r_j - d_j)}{r_j} \end{aligned}$$

Linear rank type variance:

$$var = \frac{n_0 n_1 \, \Sigma_{j=1}^{n} \, w_j^2}{n(n-1)}$$

Now consider an example where there are no tied death times

Example I Group 0: 15, 18, 19, 21, 22

Group 1: 16+, 17+, 20+, 23, 24+

A. The CMH-type logrank statistic: (using the STRATA statement)

```
Rank Statistics

TRTMT        Log-Rank     Wilcoxon

Control        2.5952       15.000
Treated       -2.5952      -15.000


Covariance Matrix for the Log-Rank Statistics

TRTMT          Control       Treated

Control        1.21712      -1.21712
Treated       -1.21712       1.21712


Test of Equality over Strata

                                 Pr >
Test      Chi-Square    DF  Chi-Square

Log-Rank      5.5338     1      0.0187
Wilcoxon      4.3269     1      0.0375
-2Log(LR)     3.1202     1      0.0773
```

B. The "linear rank" logrank statistic: (using the TEST statement)

```
Univariate Chi-Squares for the LOG RANK Test

              Test       Standard                    Pr >
Variable    Statistic    Deviation    Chi-Square    Chi-Square

TRTMT          2.5952       1.1032        5.5338        0.0187


Covariance Matrix for the LOG RANK Statistics

Variable          TRTMT

TRTMT           1.21712
```

Note that this time, the variances of the two logrank statistics are exactly the same, equal to 1.217.

If there are no tied event times, then the two versions of the test will yield identical results. The more ties we have, the more it matters which version we use.

Gehan's Generalized Wilcoxon Test

First, let's review the Wilcoxon test for uncensored data:

Denote observations from two samples by:

$$(X_1, X_2, \ldots, X_n) \text{ and } (Y_1, Y_2, \ldots, Y_m)$$

Order the combined sample and define:

$$Z_{(1)} < Z_{(2)} < \cdots < Z_{(m+n)}$$

$$R_{i1} = \text{rank of } X_i$$

$$R_1 = \sum_{i=1}^{m+n} R_{i1}$$

Reject H_0 if R_1 is too big or too small, according to

$$\frac{R_1 - E(R_1)}{\sqrt{Var(R_1)}} \sim N(0,1)$$

where

$$E(R_1) = \frac{m(m+n+1)}{2}$$

$$Var(R_1) = \frac{mn(m+n+1)}{12}$$

The **Mann-Whitney** form of the Wilcoxon is defined as:

$$U(X_i, Y_j) = U_{ij} = \begin{cases} +1 & \text{if} \quad X_i > Y_j \\ 0 & \text{if} \quad X_i = Y_j \\ -1 & \text{if} \quad X_i < Y_j \end{cases}$$

and

$$U = \sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij}.$$

There is a simple correspondence between U and R_1:

$$R_1 = m(m+n+1)/2 + U/2$$

$$\text{so} \qquad U = 2R_1 - m(m+n+1)$$

Therefore,

$$E(U) = 0$$

$$Var(U) = mn(m+n+1)/3$$

Extending Wilcoxon to censored data

The Mann-Whitney form leads to a generalization for censored data. Define

$$U(X_i, Y_j) = U_{ij} = \begin{cases} +1 & \text{if} \quad x_i > y_j \ \text{ or } \ x_i^+ \geq y_j \\ 0 & \text{if} \quad x_i = y_i \ \text{ or lower value censored} \\ -1 & \text{if} \quad x_i < y_j \ \text{ or } \ x_i \leq y_j^+ \end{cases}$$

Then define

$$W = \sum_{i=1}^{n} \sum_{j=1}^{m} U_{ij}$$

Thus, there is a contribution to W for every comparison where both observations are failures (except for ties), or where a censored observation is greater than or equal to a failure.

Looking at all possible pairs of individuals between the two treatment groups makes this a nightmare to compute by hand!

Gehan found an easier way to compute the above. First, pool the sample of $(n+m)$ observations into a single group, then compare each individual with the remaining $n+m-1$: For comparing the i-th individual with the j-th, define

$$U_{ij} = \begin{cases} +1 & \text{if} \quad t_i > t_j \ \text{ or } \ t_i^+ \geq t_j \\ -1 & \text{if} \quad t_i < t_j \ \text{ or } \ t_i \leq t_j^+ \\ 0 & \qquad otherwise \end{cases}$$

Then

$$U_i = \sum_{j=1}^{m+n} U_{ij}$$

Thus, for the i-th individual, U_i is the number of observations which are definitely less than t_i minus the number of observations that are definitely greater than t_i. We assume censorings occur after deaths, so that if $t_i = 18^+$ and $t_j = 18$, then we add 1 to U_i.

The Gehan statistic is defined as

$$\begin{aligned} U &= \sum_{i=1}^{m+n} U_i \, \mathbf{1}_{\{i \text{ in group } 0\}} \\ &= W \end{aligned}$$

U has mean 0 and variance

$$var(U) = \frac{mn}{(m+n)(m+n-1)} \sum_{i=1}^{m+n} U_i^2$$

Example from Lee:

Group 0: 15, 18, 19, 19, 20

Group 1: 16+, 18+, 20+, 23, 24+

Time	Group	U_i	U_i^2
15	0	-9	81
16^+	1	1	1
18	0	-6	36
18^+	1	2	4
19	0	-2	4
19	0	-2	4
20	0	1	1
20^+	1	5	25
23	1	4	16
24^+	1	6	36
SUM		-18	208

$$
\begin{aligned}
U &= -18 \\
Var(U) &= \frac{(5)(5)(208)}{(10)(9)} \\
&= 57.78 \\
\text{and} \quad \chi^2 &= (-18)^2/57.78 = 5.61
\end{aligned}
$$

SAS code:

```
data leedata;
  infile 'lee.dat';
  input time cens group;

proc lifetest data=leedata;
  time time*cens(0);
  strata group;
run ;
```

SAS OUTPUT: Gehans Wilcoxon test

```
Rank Statistics

TRTMT        Log-Rank     Wilcoxon

Control        2.7500       18.000
Treated       -2.7500      -18.000

Covariance Matrix for the Wilcoxon Statistics

TRTMT          Control       Treated

Control        58.4000      -58.4000
Treated       -58.4000       58.4000

Test of Equality over Strata

                                 Pr >
Test      Chi-Square    DF  Chi-Square

Log-Rank      6.9540     1      0.0084
Wilcoxon      5.5479     1      0.0185 **this is Gehan's test
-2Log(LR)     3.3444     1      0.0674
```

Notes about SAS Wilcoxon Test:

SAS calculates the Wilcoxon as $-U$ instead of U, probably so that the sign of the test statistic is consistent with the logrank.

SAS gets something slightly different for the variance, and this does not seem to depend on whether there are ties.

For example, the hypothetical dataset on p.6 without ties yields $U = -15$ and $\Sigma U_i^2 = 182$, so

$$Var(U) = \frac{(5)(5)(182)}{(10)(9)} = 50.56 \quad \text{and} \quad \chi^2 = \frac{(-15)^2}{50.56} = 4.45$$

while SAS gives the following:

```
Rank Statistics

TRTMT        Log-Rank     Wilcoxon

Control        2.5952       15.000
Treated       -2.5952      -15.000

Covariance Matrix for the Wilcoxon Statistics

TRTMT          Control       Treated

Control        52.0000      -52.0000
Treated       -52.0000       52.0000

Test of Equality over Strata
                                  Pr >
Test      Chi-Square    DF  Chi-Square

Log-Rank      5.5338     1      0.0187
Wilcoxon      4.3269     1      0.0375
-2Log(LR)     3.1202     1      0.0773
```

Obtaining the Wilcoxon test using Stata

Use the sts test statement, with the appropriate option

```
    sts test    varlist [if exp] [in range]
                [, [logrank|wilcoxon|cox] strata(varlist) detail
                   mat(matname1 matname2) notitle noshow ]

logrank, wilcoxon, and cox specify which test of equality is desired.
    logrank is the default, and cox yields a likelihood ratio test
    under a cox model.
```

Example: (leukemia data)

```
. stset remiss status

. sts test trt, wilcoxon

Wilcoxon (Breslow) test for equality of survivor functions
------------------------------------------------------------

      |  Events                         Sum of
trt   |  observed       expected         ranks
------+-------------------------------------
0     |        21          10.75           271
1     |         9          19.25          -271
------+-------------------------------------
Total |        30          30.00             0

           chi2(1) =      13.46
           Pr>chi2 =     0.0002
```

Generalized Wilcoxon
(Peto & Peto, Prentice)

Assign the following scores:

For a death at t:	$\hat{S}(t+) + \hat{S}(t-) - 1$
For a censoring at t:	$\hat{S}(t+) - 1$

The test statistic is $\Sigma(scores)$ for group 0.

Time	Group	d_j	r_j	$\hat{S}(t+)$	score w_j
15	0	1	10	0.900	0.900
16^+	1	0	9	0.900	-0.100
18	0	1	8	0.788	0.688
18^+	1	0	7	0.788	-0.212
19	0	2	6	0.525	0.313
20	0	1	4	0.394	-0.081
20^+	1	0	3	0.394	-0.606
23	1	1	2	0.197	-0.409
24^+	1	0	1	0.197	-0.803

$$\begin{aligned} \sum w_j \mathbf{1}_{\{j \text{ in group } 0\}} &= 0.900 + 0.688 + 2 * (0.313) + (-0.081) \\ &= 2.13 \end{aligned}$$

$$Var(S) = \frac{n_0 n_1 \Sigma_{j=1}^{n} w_j^2}{n(n-1)} = 0.765$$

$$\text{so} \quad Z = 2.13/0.765 = 2.433$$

The Tarone-Ware class of tests:

This general class of tests is like the logrank test, but adds weights w_j. The logrank test, Wilcoxon test, and Peto-Prentice Wilcoxon are included as special cases.

$$\chi^2_{tw} = \frac{[\Sigma_{j=1}^{K} w_j(d_{1j} - r_{1j} * d_j/r_j)]^2}{\Sigma_{l=1}^{K} \frac{w_j^2 r_{1j} r_{0j} d_j (r_j - d_j)}{r_j^2 (r_j - 1)}}$$

Test	**Weight** w_j
Logrank	$w_j = 1$
Gehan's Wilcoxon	$w_j = r_j$
Peto/Prentice	$w_j = n\widehat{S}(t_j)$
Fleming-Harrington	$w_j = [\hat{S}(t_j)]^{\alpha}$
Tarone-Ware	$w_j = \sqrt{r_j}$

Note: these weights w_j are not the same as the scores w_j we've been talking about earlier, and they apply to the CMH-type form of the test statistic rather than $\Sigma(scores)$ over a single treatment group.

Which test should we used?

CMH-type or Linear Rank?

If there are not a high proportion of ties, then it doesn't really matter since:

- The two Wilcoxons are similar to each other
- The two logrank tests are similar to each other

Note: personally, I tend to use the CMH-type test, which you get with the STRATA statement in SAS and the TEST statement in STATA.

Logrank or Wilcoxon?

- Both tests have the right Type I power for testing the null hypothesis of equal survival, $H_o : S_1(t) = S_2(t)$
- The choice of which test may therefore depend on the alternative hypothesis, which will drive the power of the test.

- The Wilcoxon is sensitive to early differences between survival, while the logrank is sensitive to later ones. This can be seen by the relative weights they assign to the test statistic:

$$\begin{aligned} \text{LOGRANK} \quad numerator &= \sum_j (o_j - e_j) \\ \text{WILCOXON} \quad numerator &= \sum_j r_j (o_j - e_j) \end{aligned}$$

- The logrank is most powerful under the assumption of proportional hazards, which implies an alternative in terms of the survival functions of $H_a : S_1(t) = [S_2(t)]^\alpha$

- The Wilcoxon has high power when the failure times are lognormally distributed, with equal variance in both groups but a different mean. It will turn out that this is the assumption of an accelerated failure time model.

- Both tests will lack power if the survival curves (or hazards) "cross". However, that does not necessarily make them invalid!

Comparison between TEST and STRATA in SAS for 2 examples:

Data from Lee (n=10):

from STRATA:

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square	
Log-Rank	6.9540	1	0.0084	
Wilcoxon	5.5479	1	0.0185	**this is Gehan's test
-2Log(LR)	3.3444	1	0.0674	

from TEST:

Univariate Chi-Squares for the WILCOXON Test

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
GROUP	1.8975	0.7508	6.3882	0.0115

Univariate Chi-Squares for the LOG RANK Test

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
GROUP	2.7500	1.0897	6.3684	0.0116

Previous example with leukemia data:

from STRATA:

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	16.7929	1	0.0001
Wilcoxon	13.4579	1	0.0002
-2Log(LR)	16.4852	1	0.0001

from TEST:

Univariate Chi-Squares for the WILCOXON Test

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
GROUP	6.6928	1.7874	14.0216	0.0002

Univariate Chi-Squares for the LOG RANK Test

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
GROUP	10.2505	2.5682	15.9305	0.0001

P-sample and stratified logrank tests

We have been discussing two sample problems. In practice, more complex settings often arise:

- There are more than two treatments or groups, and the question of interest is whether the groups differ from each other.

- We are interested in a comparison between two groups, but we wish to adjust for another factor that may confound the analysis

- We want to adjust for lots of covariates.

We will first talk about comparing the survival distributions between more than 2 groups, and then about adjusting for other covariates.

P-sample logrank

Suppose we observe data from P different groups, and the data from group p ($p = 1, ..., P$) are:

$$(X_{p1}, \delta_{p1}) \ldots (X_{pn_p}, \delta_{pn_p})$$

We now construct a $(P \times 2)$ table at each of the K distinct death times, and compare the death rates between the P groups, conditional on the number at risk.

Let $t_1, ...t_K$ represent the K ordered, distinct death times. At the j-th death time, we have the following table:

	Die/Fail		
Group	Yes	No	Total
1	d_{1j}	$r_{1l} - d_{1j}$	r_{1j}
.	.	.	.
P	d_{Pj}	$r_{Pj} - d_{Pj}$	r_{Pj}
Total	d_j	$r_j - d_j$	r_j

where d_{pj} is the number of deaths in group p at the j-th death time, and r_{pj} is the number at risk at that time.

The tables are then combined using the CMH approach.

If we were just focusing on this one table, then a $\chi^2_{(P-1)}$ test statistic could be constructed through a comparison of "o"s and "e"s, like before.

Example: Toxicity in a clinical trial with 3 treatments

```
TABLE OF GROUP BY TOXICITY

GROUP     TOXICITY

Frequency|
Row Pct  |no      |yes     |  Total
---------+--------+--------+
       1 |     42 |      8 |     50
         |  84.00 |  16.00 |
---------+--------+--------+
       2 |     48 |      2 |     50
         |  96.00 |   4.00 |
---------+--------+--------+
       3 |     38 |     12 |     50
         |  76.00 |  24.00 |
---------+--------+--------+
Total         128       22      150


STATISTICS FOR TABLE OF GROUP BY TOXICITY

Statistic                     DF     Value        Prob
------------------------------------------------------
Chi-Square                     2     8.097       0.017
Likelihood Ratio Chi-Square    2     9.196       0.010
Mantel-Haenszel Chi-Square     1     1.270       0.260


  Cochran-Mantel-Haenszel Statistics (Based on Table Scores)

Statistic   Alternative Hypothesis    DF    Value     Prob
-------------------------------------------------------------
   1        Nonzero Correlation        1    1.270    0.260
   2        Row Mean Scores Differ     2    8.043    0.018
   3        General Association        2    8.043    0.018
```

Formal Calculations:

Let $\mathbf{O}_j = (d_{1j}, ...d_{(P-1)j})^T$ be a vector of the observed number of failures in groups 1 to $(P-1)$, respectively, at the j-th death time. Given the risk sets r_{1j}, ... r_{Pj}, and the fact that there are d_j deaths, then $\mathbf{O}_j$ has a distribution like a multivariate version of the Hypergeometric. $\mathbf{O}_j$ has mean:

$$\mathbf{E}_j = (\frac{d_j\, r_{1j}}{r_j},\ ...\ , \frac{d_j\, r_{(P-1)j}}{r_j})^T$$

and variance covariance matrix:

$$\mathbf{V}_j = \begin{pmatrix} v_{11j} & v_{12j} & ... & v_{1(P-1)j} \\ & v_{22j} & ... & v_{2(P-1)j} \\ ... & & ... & ... \\ & & & v_{(P-1)(P-1)j} \end{pmatrix}$$

where the ℓ-th diagonal element is:

$$v_{\ell\ell j} = r_{\ell j}(r_j - r_{\ell j})d_j(r_j - d_j)/[r_j^2(r_j - 1)]$$

and the ℓm-th off-diagonal element is:

$$v_{\ell m j} = r_{\ell j} r_{mj} d_j(r_j - d_j)/[r_j^2(r_j - 1)]$$

The resulting χ^2 test for a single $(P \times 1)$ table would have (P-1) degrees and is constructed as follows:

$$(\mathbf{O}_j - \mathbf{E}_j)^T \, \mathbf{V}_j^{-1} \, (\mathbf{O}_j - \mathbf{E}_j)$$

Generalizing to K tables

Analogous to what we did for the two sample logrank, we replace the $\mathbf{O}_j$, $\mathbf{E}_j$ and $\mathbf{V}_j$ with the sums over the K distinct death times. That is, let $\mathbf{O} = \Sigma_{j=1}^{k} \mathbf{O}_j$, $\mathbf{E} = \Sigma_{j=1}^{k} \mathbf{E}_j$, and $\mathbf{V} = \Sigma_{j=1}^{k} \mathbf{V}_j$. Then, the test statistic is:

$$(\mathbf{O} - \mathbf{E})^T \, \mathbf{V}^{-1} \, (\mathbf{O} - \mathbf{E})$$

Example:
Time taken to finish a test with 3 different noise distractions. All tests were stopped after 12 minutes.

Noise Level		
Group 1	Group 2	Group 3
9.0	10.0	12.0
9.5	12.0	12^+
9.0	12^+	12^+
8.5	11.0	12^+
10.0	12.0	12^+
10.5	10.5	12^+

Lets start the calculations ...

Observed data table

Ordered	Group 1		Group 2		Group 3		Combined	
Times	d_{1j}	r_{1j}	d_{2j}	r_{2j}	d_{3j}	r_{3j}	d_j	r_j
8.5	1	6	0	6	0	6		
9.0	2	5	0	6	0	6		
9.5	1	3	0	6	0	6		
10.0	1	2	1	6	0	6		
10.5	1	1	1	5	0	6		
11.0	0	0	1	4	0	6		
12.0	0	0	2	3	1	6		

Expected table

Ordered	Group 1		Group 2		Group 3		Combined	
Times	o_{1j}	e_{1j}	o_{2j}	e_{2j}	o_{3j}	e_{3j}	o_j	e_j
8.5								
9.0								
9.5								
10.0								
10.5								
11.0								
12.0								

Doing the P-sample test by hand is cumbersome ...

Luckily, most statistical packages will do it for you!

P-sample logrank in Stata

```
.sts graph, by(group)

.sts test group, logrank

Log-rank test for equality of survivor functions
------------------------------------------------

      |  Events
group |  observed       expected
------+--------------------------
1     |         6           1.57
2     |         5           4.53
3     |         1           5.90
------+--------------------------
Total |        12          12.00

           chi2(2) =      20.38
           Pr>chi2 =     0.0000


. sts test group, wilcoxon

Wilcoxon (Breslow) test for equality of survivor functions
----------------------------------------------------------

      |  Events                         Sum of
group |  observed       expected         ranks
------+---------------------------------------
1     |         6           1.57            68
2     |         5           4.53            -5
3     |         1           5.90           -63
------+---------------------------------------
Total |        12          12.00             0

           chi2(2) =      18.33
           Pr>chi2 =     0.0001
```

SAS program for P-sample logrank

```
Title 'Testing with noise example';
data noise;
   input testtime finish group;
   cards;
9       1    1
9.5     1    1
9.0     1    1
8.5     1    1
10      1    1
10.5    1    1
10.0    1    2
12      1    2
12      0    2
11      1    2
12      1    2
10.5    1    2
12      1    3
12      0    3
12      0    3
12      0    3
12      0    3
12      0    3
;

proc lifetest data=noise;
  time testtime*finish(0);
  strata group;
run;
```

Testing Homogeneity of Survival Curves over Strata

Time Variable TESTTIME

Rank Statistics

GROUP	Log-Rank	Wilcoxon
1	4.4261	68.000
2	0.4703	-5.000
3	-4.8964	-63.000

Covariance Matrix for the Log-Rank Statistics

GROUP	1	2	3
1	1.13644	-0.56191	-0.57454
2	-0.56191	2.52446	-1.96255
3	-0.57454	-1.96255	2.53709

Covariance Matrix for the Wilcoxon Statistics

GROUP	1	2	3
1	284.808	-141.495	-143.313
2	-141.495	466.502	-325.007
3	-143.313	-325.007	468.320

Test of Equality over Strata

Test	Chi-Square	DF	Pr > Chi-Square
Log-Rank	20.3844	2	0.0001
Wilcoxon	18.3265	2	0.0001
-2Log(LR)	5.5470	2	0.0624

Note: do not use TEST in SAS PROC LIFETEST if you want a P-sample logrank. TEST will interpret the group variable as a measured covariate (i.e., either ordinal or continuous).

In other words, you will get a *trend* test with only 1 degree of freedom, rather than a P-sample test with (p-1) df.

For example, here's what we get if we use the TEST statement on the noise example:

```
proc lifetest data=noise;
  time testtime*finish(0);
  test group;
run;
```

SAS OUTPUT:

```
Univariate Chi-Squares for the LOG RANK Test
              Test       Standard                    Pr >
Variable    Statistic    Deviation    Chi-Square    Chi-Square

GROUP          9.3224       2.2846      16.6503       0.0001

Covariance Matrix for the LOG RANK Statistics
Variable         GROUP

GROUP          5.21957

Forward Stepwise Sequence of Chi-Squares for the LOG RANK Test
                                  Pr >      Chi-Square      Pr >
Variable     DF   Chi-Square   Chi-Square    Increment   Increment

GROUP         1     16.6503      0.0001       16.6503      0.0001
```

The Stratified Logrank

Sometimes, even though we are interested in comparing two groups (or maybe P) groups, we know there are other factors that also affect the outcome. It would be useful to adjust for these other factors in some way.

Example: For the nursing home data, a logrank test comparing length of stay for those under and over 85 years of age suggests a significant difference (p=0.03).

However, we know that gender has a strong association with length of stay, and also age. Hence, it would be a good idea to STRATIFY the analysis by gender when trying to assess the age effect.

A **stratified logrank** allows one to compare groups, but allows the shapes of the hazards of the different groups to differ across strata. It makes the assumption that the group 1 vs group 2 hazard ratio is constant across strata.

In other words: $\frac{\lambda_{1s}(t)}{\lambda_{2s}(t)} = \theta$ where θ is constant over the strata ($s = 1, ..., S$).

This method of adjusting for other variables is not as flexible as that based on a modelling approach.

General setup for the stratified logrank:

Suppose we want to assess the association between survival and a factor (call this X) that has two different levels. Suppose however, that we want to stratify by a second factor, that has S different levels.

First, divide the data into S separate groups. Within group s ($s = 1, ..., S$), proceed as though you were constructing the logrank to assess the association between survival and the variable X. That is, let $t_{1s}, ..., t_{K_s s}$ represent the K_s ordered, distinct death times in the s-th group.

At the j-th death time in group s, we have the following table:

	Die/Fail		
X	Yes	No	Total
1	d_{s1j}	$r_{s1j} - d_{s1j}$	r_{s1j}
2	d_{s2j}	$r_{s2j} - d_{s2j}$	r_{s2j}
Total	d_{sj}	$r_{sj} - d_{sj}$	r_{sj}

Let O_s be the sum of the "o"s obtained by applying the logrank calculations in the usual way to the data from group s. Similarly, let E_s be the sum of the "e"s, and V_s be the sum of the "v"s.

The **stratified logrank** is

$$Z = \frac{\Sigma_{s=1}^{S}(O_s - E_s)}{\sqrt{\Sigma_{s=1}^{S}(V_s)}}$$

Stratified logrank using Stata:

```
. use nurshome

. gen age1=0

. replace age1=1 if age>85

. sts test age1, strata(gender)

         failure _d:  cens
   analysis time _t:  los


Stratified log-rank test for equality of survivor functions
------------------------------------------------------------

      |  Events
age1  |  observed    expected(*)
------+--------------------------
0     |       795         764.36
1     |       474         504.64
------+--------------------------
Total |      1269        1269.00

(*) sum over calculations within gender

            chi2(1) =       3.22
            Pr>chi2 =     0.0728
```

Stratified logrank using SAS:

```
data pop1;
  set pop;
  age1=0;
  if age >85 then age1=1;

proc lifetest data=pop1 outsurv=survres;
  time stay*censor(1);
  test age1;
  strata gender;
```

RESULTS (just the logrank part .... you can also do a stratified Wilcoxon)

The LIFETEST Procedure

Rank Tests for the Association of LSTAY with Covariates
Pooled over Strata

Univariate Chi-Squares for the LOG RANK Test

Variable	Test Statistic	Standard Deviation	Chi-Square	Pr > Chi-Square
AGE1	29.1508	17.1941	2.8744	0.0900

Covariance Matrix for the LOG RANK Statistics

Variable	AGE1
AGE1	295.636

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

Variable	DF	Chi-Square	Pr > Chi-Square	Chi-Square Increment	Pr > Increment
AGE1	1	2.8744	0.0900	2.8744	0.0900

Modeling of Survival Data

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

- **Proportional Hazards (PH) models**

  $$\lambda(t; \mathbf{Z}) = \lambda_0(t)\Psi(\mathbf{Z})$$

  Most commonly, we write the second term as:

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

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

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

- **Accelerated Failure Time (AFT) models**

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

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

  – exponential, Weibull, Gamma

  – lognormal

Covariates:

In general, $\mathbf{Z}$ is a *vector* of covariates of interest.

$\mathbf{Z}$ may include:

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

Discrete Covariates:

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

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

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

Interactions:

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

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

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

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

- What does the term $\lambda_0(t)$ mean?
- What's "proportional" about the PH model?
- How do we estimate the parameters in the model?
- How do we interpret the estimated values?
- How can we construct tests of whether the covariates have a significant effect on the distribution of survival times?
- How do these tests compare to the logrank test or the Wilcoxon test?

The Cox Proportional Hazards model

$$\lambda(t; \mathbf{Z}) = \lambda_0(t) \exp(\boldsymbol{\beta}\mathbf{Z})$$

This is the most common model used for survival data.

Why?

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

References: Collett, Chapter 3*
Lee, Chapter 10*
Hosmer & Lemeshow, Chapters 3-7
Allison, Chapter 5
Cox and Oakes, Chapter 7
Kleinbaum, Chapter 3
Klein and Moeschberger, Chapters 8 & 9
Kalbfleisch and Prentice

Note: some books (like Collett and H & L) use $h(t; \mathbf{X})$ as their standard notation for the hazard instead of $\lambda(t; \mathbf{Z})$, and $H(t)$ for the cumulative hazard instead of $\Lambda(t)$.

Why do we call it proportional hazards?

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

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

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

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

ϕ is referred to as the **hazard ratio**.

What is the interpretation of β here?

The Baseline Hazard Function

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

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

The general form is:

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

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

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

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

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

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

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

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

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

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

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

Questions:

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

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

How do we estimate the model parameters?

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

Suppose T follows a PH model:

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

Now consider $T^* = g(T)$, where g is a monotonic increasing function. We can show that T^* also follows the PH model, with the same multiplier, $e^{\boldsymbol{\beta}\mathbf{Z}}$.

Therefore, when we consider likelihood methods for estimating the model parameters, we only have to worry about the ranks of the survival times.

Likelihood Estimation for the PH Model

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

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

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

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

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

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

For now, assume there are no tied death times.

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

More about risk sets:

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

What is the partial likelihood?

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

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

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

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

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

Another derivation:

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

- **Individual is censored at X_i:**

$$L_i(\boldsymbol{\beta}) = S(X_i) = \exp[-\int_0^{X_i} \lambda_i(u)du]$$

- **Individual fails at X_i:**

$$L_i(\boldsymbol{\beta}) = S(X_i)\lambda_i(X_i) = \lambda_i(X_i)\exp[-\int_0^{X_i} \lambda_i(u)du]$$

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

This means we get a total likelihood of:

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

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

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

$$L(\boldsymbol{\beta}) \;=\; \prod_{i=1}^{n}\left[\frac{\lambda_i(X_i)}{\Sigma_{j\in\mathcal{R}(X_i)}\,\lambda_i(X_i)}\right]^{\delta_i}\;\left[\sum_{j\in\mathcal{R}(X_i)}\lambda_i(X_i)\right]^{\delta_i}\;\exp[-\int_0^{X_i}\lambda_i(u)du]$$

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

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

$$\begin{aligned} L(\boldsymbol{\beta}) \;&=\; \prod_{i=1}^{n}\left[\frac{\lambda_i(X_i)}{\Sigma_{j\in\mathcal{R}(X_i)}\,\lambda_i(X_i)}\right]^{\delta_i} \\ &=\; \prod_{i=1}^{n}\left[\frac{\lambda_0(X_i)\exp(\boldsymbol{\beta}\mathbf{Z}_i)}{\Sigma_{j\in\mathcal{R}(X_i)}\,\lambda_0(X_i)\exp(\boldsymbol{\beta}\mathbf{Z}_j)}\right]^{\delta_i} \\ &=\; \prod_{i=1}^{n}\left[\frac{\exp(\boldsymbol{\beta}\mathbf{Z}_i)}{\Sigma_{j\in\mathcal{R}(X_i)}\exp(\boldsymbol{\beta}\mathbf{Z}_j)}\right]^{\delta_i} \end{aligned}$$

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

A simple example:

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

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

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

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

Notes on the partial likelihood:

$$L(\boldsymbol{\beta}) = \prod_{j=1}^{n} \left[\frac{e^{\boldsymbol{\beta}\mathbf{Z}_j}}{\sum_{\ell \in \mathcal{R}(X_j)} e^{\boldsymbol{\beta}\mathbf{Z}_\ell}} \right]^{\delta_j}$$

$$= \prod_{j=1}^{K} \frac{e^{\boldsymbol{\beta}\mathbf{Z}_j}}{\sum_{\ell \in \mathcal{R}(\tau_j)} e^{\boldsymbol{\beta}\mathbf{Z}_\ell}}$$

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

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

Partial Likelihood inference

Inference can be conducted by treating the partial likelihood as though it satisfied all the regular likelihood properties.

The **log-partial likelihood** is:

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

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

Suppose there is only one covariate ($\boldsymbol{\beta}$ is one-dimensional):

The **partial likelihood score equations** are:

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

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

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

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

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

Analogous to standard likelihood theory, it can be shown (though not easily) that

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

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

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

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

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

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

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

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

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

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

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

What does β represent in this case?

Notes

- The "observed" information matrix is generally used because in practice, people find it has better properties. Also, the "expected" is very hard to calculate.

- There is a nice analogy with the score and information matrices from more standard regression problems, except that here we are summing over observed death times, rather than individuals.

- Newton Raphson is used by many of the computer packages to solve the partial likelihood equations.

Fitting Cox PH model with Stata

Uses the "`stcox`" command.

First, try typing "`help stcox`"

```
-------------------------------------------------------------------------------
help for stcox
-------------------------------------------------------------------------------

Estimate Cox proportional hazards model
---------------------------------------

stcox [varlist] [if exp] [in range]
          [, nohr strata(varnames) robust cluster(varname) noadjust
          mgale(newvar) esr(newvars)
          schoenfeld(newvar) scaledsch(newvar)
          basehazard(newvar) basechazard(newvar) basesurv(newvar)
          {breslow | efron | exactm | exactp} cmd estimate noshow
          offset level(#) maximize-options ]

stphtest [, km log rank time(varname) plot(varname) detail
          graph-options ksm-options]

stcox is for use with survival-time data; see help st.  You must
have stset your data before using this command; see help stset.

Description
-----------
stcox estimates maximum-likelihood proportional hazards models on st data.

Options  (many more!)
-------
nohr reports the estimated coefficients rather than hazard ratios; i.e.,
    b rather than exp(b).  Standard errors and confidence intervals are
    similarly transformed.  This option affects how results are displayed,
    not how they are estimated.
```

Ex. Leukemia Data

```
. stcox trt

Iteration 0:   log likelihood =  -93.98505
Iteration 1:   log likelihood = -86.385606
Iteration 2:   log likelihood = -86.379623
Iteration 3:   log likelihood = -86.379622
Refining estimates:
Iteration 0:   log likelihood = -86.379622

Cox regression -- Breslow method for ties

No. of subjects =           42                     Number of obs   =        42
No. of failures =           30
Time at risk    =          541
                                                   LR chi2(1)      =     15.21
Log likelihood  =   -86.379622                     Prob > chi2     =    0.0001

------------------------------------------------------------------------------
      _t |
      _d | Haz. Ratio   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     trt |   .2210887   .0905501     -3.685   0.000       .0990706    .4933877
------------------------------------------------------------------------------

. stcox trt , nohr

(same iterations for log-likelihood)

Cox regression -- Breslow method for ties

No. of subjects =           42                     Number of obs   =        42
No. of failures =           30
Time at risk    =          541
                                                   LR chi2(1)      =     15.21
Log likelihood  =   -86.379622                     Prob > chi2     =    0.0001

------------------------------------------------------------------------------
      _t |
      _d |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     trt |  -1.509191   .4095644     -3.685   0.000      -2.311923   -.7064599
------------------------------------------------------------------------------
```

Fitting PH models in SAS - PROC PHREG

Ex. Leukemia data

```
Title 'Cox and Oakes example';
data leukemia;
    input weeks remiss trtmt;
    cards;
6     0     1
6     1     1
6     1     1
6     1     1        /* data for 6MP group */
7     1     1
9     0     1
etc
1     1     0
1     1     0        /* data for placebo group */
2     1     0
2     1     0
etc
;

proc phreg data=leukemia;
  model weeks*remiss(0)=trtmt;
  title 'Cox PH Model for leukemia data';
run;
```

PROC PHREG Output:

The PHREG Procedure

Data Set: WORK.LEUKEM
Dependent Variable: FAILTIME Time to Relapse
Censoring Variable: FAIL
Censoring Value(s): 0
Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

Total	Event	Censored	Percent Censored
42	30	12	28.57

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	187.970	172.759	15.211 with 1 DF (p=0.0001)
Score	.	.	15.931 with 1 DF (p=0.0001)
Wald	.	.	13.578 with 1 DF (p=0.0002)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
TRTMT	1	-1.509191	0.40956	13.57826	0.0002	0.221

Fitting PH models in S-plus: `coxph` function

Here are some of the data in leuk.dat:

```
t f x
1 1 0
1 1 0
2 1 0
2 1 0
3 1 0
 ...
19 0 1
20 0 1
22 1 1
23 1 1
25 0 1
32 0 1
32 0 1
34 0 1
35 0 1
```

```
leuk_read.table("leuk.dat",header=T)

#specify Breslow handling of ties
print(coxph(Surv(t,f) ~ x, leuk, method="breslow"))

#specify Efron handling of ties (default)
print(coxph(Surv(t,f) ~ x, leuk))
```

coxph Output:

```
Call:
coxph(formula = Surv(t, f) ~ x, data = leuk, method = "breslow")


   coef exp(coef) se(coef)     z       p
x -1.51     0.221     0.41 -3.68 0.00023

Likelihood ratio test=15.2  on 1 df, p=0.0000961  n= 42
```

```
Call:
coxph(formula = Surv(t, f) ~ x, data = leuk)


   coef exp(coef) se(coef)     z       p
x -1.57     0.208    0.412 -3.81 0.00014

Likelihood ratio test=16.4  on 1 df, p=0.0000526  n= 42
```

Compare this with the logrank test from PROC LIFETEST (Using the "TEST" statement)

```
The LIFETEST Procedure

Rank Tests for the Association of FAILTIME with Covariates
Pooled over Strata

Univariate Chi-Squares for the LOG RANK Test

              Test       Standard                    Pr >
Variable    Statistic    Deviation    Chi-Square    Chi-Square

TRTMT         10.2505       2.5682       15.9305        0.0001
```

Notes:

- **The logrank test=score test from `Proc phreg`!** In general, the score test would be for *all* of the variables in the model, but in this case, we have only "`trtmt`".
- Stata does not provide a score test in its output from the Cox model. However, the `stcox` command with the `breslow` option for ties yields the same LR test as the CMH-version logrank test from the `sts test, cox` command.

More Notes:

- The Cox Proportional hazards model has the advantage over a simple logrank test of giving us an estimate of the "risk ratio" (i.e., $\phi = \lambda_1(t)/\lambda_0(t)$). This is more informative than just a test statistic, and we can also form confidence intervals for the risk ratio.

- In this case, $\hat{\phi} = 0.221$, which can be interpreted to mean that the hazard for relapse among patients treated with 6-MP is less than 25% of that for placebo patients.

- From the STS LIST command in Stata or PROC LIFETEST in SAS, we were able to get estimates of the entire survival distribution $\hat{S}(t)$ for each treatment group; we can't immediately get this from our Cox model without further assumptions. **Why not?**

Adjustments for ties

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

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

(2) **Peto-Breslow method**

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

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

(5) **Exact marginal method** (stata)

Some notation:

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

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

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

The **partial likelihood** is:

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

where

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

What does this all mean??!!

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

Simple Example (with ties)

Group 0: $4^+, 6, 8^+, 9, 10^+ \implies Z_i = 0$
Group 1: $3, 5, 5^+, 6, 8^+ \implies Z_i = 1$

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

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

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

(2) Breslow method: (default)

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

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

Justification:

Suppose individuals 1 and 2 fail from $\{1, 2, 3, 4\}$ at time τ_j. Let $\phi(i)$ be the hazard ratio for individual i (compared to baseline).

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

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

(3) Efron's (1977) method:

Efron suggested an even closer approximation to the discrete likelihood:

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

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

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

(4) Exact method (Kalbfleisch and Prentice):

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

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

All possible orderings of the tied events are calculated, and the probabilities of each are summed.

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

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

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

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

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

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

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

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

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

Example: The fecundability study

Women who had recently given birth (or had tried to get pregnant for at least a year) were asked to recall how long it took them to become pregnant, and whether or not they smoked during that time. The outcome of interest is time to pregnancy (measured in menstrual cycles).

```
data fecund;
  input   smoke      cycle     status    count;
  cards;
0          1          1        198
0          2          1        107
0          3          1         55
0          4          1         38
0          5          1         18
0          6          1         22
.....................................

1         10          1          1
1         11          1          1
1         12          1          3
1         12          0          7
;

proc phreg;
  model cycle*status(0) = smoke /ties=breslow;   /* default */
  freq count;

proc phreg;
  model cycle*status(0) = smoke /ties=discrete;
  freq count;

proc phreg;
  model cycle*status(0) = smoke /ties=exact;
  freq count;

proc phreg;
  model cycle*status(0) = smoke /ties=efron;
  freq count;
```

SAS Output for Fecundability study: Accounting for Ties

```
*************************************************************************
Ties Handling: BRESLOW

                  Parameter   Standard       Wald         Pr >      Risk
Variable    DF    Estimate     Error     Chi-Square    Chi-Square   Ratio
 SMOKE      1    -0.329054    0.11412       8.31390      0.0039     0.720


*************************************************************************
Ties Handling: DISCRETE

                  Parameter   Standard       Wald         Pr >      Risk
Variable    DF    Estimate     Error     Chi-Square    Chi-Square   Ratio
 SMOKE      1    -0.461246    0.13248      12.12116      0.0005     0.630


*************************************************************************
Ties Handling: EXACT

                  Parameter   Standard       Wald         Pr >      Risk
Variable    DF    Estimate     Error     Chi-Square    Chi-Square   Ratio
 SMOKE      1    -0.391548    0.11450      11.69359      0.0006     0.676


*************************************************************************
Ties Handling: EFRON

                  Parameter   Standard       Wald         Pr >      Risk
Variable    DF    Estimate     Error     Chi-Square    Chi-Square   Ratio
 SMOKE      1    -0.387793    0.11402      11.56743      0.0007     0.679

*************************************************************************
```

For this particular dataset, does it seem like it would be important to consider the effect of tied failure times? Which method would be best?

Stata Commands for PH Model with Ties:

Stata also offers four options for adjustments with tied data:

- `breslow` (default)
- `efron`
- `exactp` (same as the "discrete" option in SAS)
- `exactm` - an exact marginal likelihood calculation (different than the "exact" option in SAS)

Fecundability Data Example:

```
. stcox smoker, efron nohr

        failure _d:  status
  analysis time _t:  cycle

Iteration 0:   log likelihood = -3113.5313
Iteration 1:   log likelihood = -3107.3102
Iteration 2:   log likelihood = -3107.2464
Iteration 3:   log likelihood = -3107.2464
Refining estimates:
Iteration 0:   log likelihood = -3107.2464
Cox regression -- Efron method for ties

No. of subjects =          586                     Number of obs   =       586
No. of failures =          567
Time at risk    =         1844
                                                   LR chi2(1)      =     12.57
Log likelihood  =   -3107.2464                     Prob > chi2     =    0.0004

------------------------------------------------------------------------------
      _t |
      _d |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
  smoker |  -.3877931   .1140202     -3.401   0.001      -.6112685   -.1643177
------------------------------------------------------------------------------
```

A special case: the two-sample problem

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

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

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

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

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

Using the proportional hazards formulation, we have

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

Group 0 hazard: $\lambda_0(t)$

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

The log-partial likelihood is:

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

Taking the derivative with respect to β, we get:

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

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

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

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

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

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

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

Why? Under $H_0 : \beta = 0$, $\bar{Z}_j$ is simply the number of individuals from group 1 in the risk set at time τ_j (call this r_{1j}), divided by the total number in the risk set at that time (call this r_j). Thus, $\bar{Z}_j$ approximates the probability that given there is a death at τ_j, it is from group 1.

Thus, the score statistic is of the form:

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

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

In SAS or Stata:

discrete/exactp → Mantel-Haenszel logrank test

breslow → linear rank version of the logrank test

I already showed you the equivalence of the linear rank logrank test and the Breslow (default) Cox PH model in SAS (p.24-25)

Here is the output from SAS for the leukemia data using the **method=discrete** option:

```
Logrank test with proc lifetest - strata statement

Test of Equality over Strata
                                 Pr >
Test      Chi-Square    DF  Chi-Square

Log-Rank     16.7929     1      0.0001
Wilcoxon     13.4579     1      0.0002
-2Log(LR)    16.4852     1      0.0001
```

```
The PHREG Procedure

Data Set: WORK.LEUKEM
Dependent Variable: FAILTIME  Time to Relapse
Censoring Variable: FAIL
Censoring Value(s): 0
Ties Handling: DISCRETE

                Testing Global Null Hypothesis: BETA=0

               Without        With
Criterion    Covariates    Covariates    Model Chi-Square

-2 LOG L        165.339       149.086      16.252 with 1 DF (p=0.0001)
Score              .             .         16.793 with 1 DF (p=0.0001)
Wald               .             .         14.132 with 1 DF (p=0.0002)
```

More on the Cox PH model

I. Confidence intervals and hypothesis tests

- **Two methods for confidence intervals**
- **Wald tests and likelihood ratio tests**
- **Interpretation of parameter estimates**
- **An example with real data from an AIDS clinical trial**

II. Predicted survival under proportional hazards

III. Predicted medians and P-year survival

I. Constructing Confidence intervals and tests for the Hazard Ratio (see H & L 4.2, Collett 3.4):

Many software packages provide estimates of β, but the hazard ratio HR= $\exp(\beta)$ is usually the parameter of interest.

We can use the delta method to get standard errors for $\exp(\hat{\beta})$:

$$Var(\widehat{HR}) \;=\; Var(\exp(\hat{\beta})) = \exp(2\hat{\beta})Var(\hat{\beta})$$

Constructing confidence intervals for $\exp(\beta)$
Two options: (assuming that β **is a scalar)**

I. Using $se(\exp\hat{\beta})$ obtained above via the delta method as $se(\exp\hat{\beta}) = \sqrt{[Var(\exp(\hat{\beta}))]}$, calculate the endpoints as:

$$[L, U] \;=\; [\widehat{OR} - 1.96\, se(\widehat{OR}), \widehat{OR} + 1.96\, se(\widehat{OR})]$$

II. Form a confidence interval for $\hat{\beta}$, and then exponentiate the endpoints.

$$[L, U] \;=\; [e^{\hat{\beta}-1.96se(\hat{\beta})}, e^{\hat{\beta}+1.96se(\hat{\beta})}]$$

Which approach do you think would be the most preferable?

Hypothesis Tests:

For each covariate of interest, the null hypothesis is

$$H_o : HR_j = 1 \Leftrightarrow \beta_j = 0$$

A Wald test[2] of the above hypothesis is constructed as:

$$Z = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \qquad \text{or} \qquad \chi^2 = \left(\frac{\hat{\beta}_j}{se(\hat{\beta}_j)}\right)^2$$

This test for $\beta_j = 0$ assumes that all other terms in the model are held fixed.

Note: if we have a factor A with a levels, then we would need to construct a χ^2 test with $(a-1)$ df, using a test statistic based on a quadratic form:

$$\chi^2_{(a-1)} = \widehat{\boldsymbol{\beta}}'_A Var(\widehat{\boldsymbol{\beta}}_A)^{-1} \widehat{\boldsymbol{\beta}}_A$$

where $\boldsymbol{\beta}_A = (\beta_2, ..., \beta_a)'$ are the $(a-1)$ coefficients corresponding to $Z_2, ..., Z_a$ (or $Z_1, ..., Z_{a-1}$, depending on the reference group).

[2]The first follows a normal distribution, and the second follows a χ^2 with 1 df. STATA gives the Z statistic, while SAS gives the χ^2_1 test statistic (the p-values are also given, and don't depend on which form, Z or χ^2, is provided)

Likelihood Ratio Tests:

Suppose there are $(p+q)$ explanatory variables measured:

$$Z_1, \ldots, Z_p, Z_{p+1}, \ldots, Z_{p+q}$$

and proportional hazards are assumed.

Consider the following models:

- **Model 1:** (contains only the first p covariates)

$$\frac{\lambda_i(t, \mathbf{Z})}{\lambda_0(t)} = \exp(\beta_1 Z_1 + \cdots + \beta_p Z_p)$$

- **Model 2:** (contains all $(p+q)$ covariates)

$$\frac{\lambda_i(t, \mathbf{Z})}{\lambda_0(t)} = \exp(\beta_1 Z_1 + \cdots + \beta_{p+q} Z_{p+q})$$

These are *nested* models. For such nested models, we can construct a **likelihood ratio** test of

$$H_0 : \beta_{p+1} = \cdots = \beta_{p+q} = 0$$

as:

$$\chi^2_{LR} \;=\; -2\left[\log(\hat{L}(1)) - \log(\hat{L}(2))\right]$$

Under H_o, this test statistic is approximately distributed as χ^2 with q df.

Some examples using the Stata `stcox` command:

Model 1:

```
. use mac

. stset mactime macstat

. stcox karnof rif clari, nohr

        failure _d:  macstat
  analysis time _t:  mactime

Cox regression -- Breslow method for ties

No. of subjects =         1151              Number of obs    =      1151
No. of failures =          121
Time at risk    =       489509
                                            LR chi2(3)       =     32.01
Log likelihood  =   -754.52813              Prob > chi2      =    0.0000

------------------------------------------------------------------------------
      _t |
      _d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
  karnof |  -.0448295   .0106355    -4.215  0.000    -.0656747   -.0239843
     rif |   .8723819   .2369497     3.682  0.000     .4079691    1.336795
   clari |   .2760775   .2580215     1.070  0.285    -.2296354    .7817903
------------------------------------------------------------------------------
```

Model 2:

```
. stcox karnof rif clari cd4, nohr

        failure _d:  macstat
  analysis time _t:  mactime


Cox regression -- Breslow method for ties

No. of subjects =         1151                 Number of obs   =      1151
No. of failures =          121
Time at risk    =       489509
                                               LR chi2(4)      =     63.74
Log likelihood  =   -738.66225                 Prob > chi2     =    0.0000

------------------------------------------------------------------------------
      _t |
      _d |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
  karnof |  -.0368538   .0106652    -3.456   0.001    -.0577572   -.0159503
     rif |    .880338   .2371111     3.713   0.000     .4156089    1.345067
   clari |   .2530205   .2583478     0.979   0.327     -.253332    .7593729
     cd4 |  -.0183553   .0036839    -4.983   0.000    -.0255757   -.0111349
------------------------------------------------------------------------------
```

Notes:

- If we omit the `nohr` option, we will get the estimated hazard ratio along with 95% confidence intervals using Method II (i.e., forming a CI for the log HR (beta), and then exponentiating the bounds)

```
------------------------------------------------------------------------------
      _t |
      _d | Haz. Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
  karnof |   .9638171   .0102793   -3.456   0.001     .9438791    .9841762
     rif |   2.411715   .5718442    3.713   0.000     1.515293    3.838444
   clari |    1.28791   .3327287    0.979   0.327     .7762102    2.136936
     cd4 |   .9818121   .0036169   -4.983   0.000     .9747486    .9889269
------------------------------------------------------------------------------
```

- We can also compute the hazard ratio ourselves, by exponentiating the coefficients:

$$HR_{cd4} \;=\; \exp(-0.01835) = 0.98$$

Why is this HR so close to 1, and yet still highly significant?

What is the interpretation of this HR?

- The likelihood ratio test for the effect of CD4 is twice the difference in minus log-likelihoods between the two models:

$$\chi^2_{LR} \;=\; 2 * (754.533 - (738.66)) = 31.74$$

How does this test statistic compare to the Wald χ^2 test?

- In the mac study, there were three treatment arms (rif, clari, and the rif+clari combination). Because we have only included the `rif` and `clari` effects in the model, the combination therapy is the "reference" group.
- We can conduct an overall test of treatment using the `test` command in Stata:

```
. test rif clari

 ( 1)  rif = 0.0
 ( 2)  clari = 0.0

           chi2(  2) =   17.01
         Prob > chi2 =    0.0002
```

  for a 2 df Wald chi-square test of whether both treatment coefficients are equal to 0. This `test` command can be used to conduct an overall test for any number of effects.

- The `test` command can also be used to test whether there is a difference between the `rif` and `clari` treatment arms:

```
.  test rif=clari

 ( 1)  rif - clari = 0.0

           chi2(  1) =    8.76
         Prob > chi2 =    0.0031
```

Some examples using SAS PROC PHREG

```
proc phreg data=alloi;
  model dthtime*dthstat(0)=mlogrna cd4grp1 cd4grp2 combther
        / risklimits;
  cd4level: test cd4grp1, cd4grp2;
  title1 'Proportional hazards regression model for time to Death';
  title2 'Baseline viral load and CD4 predictors';

proc phreg data=alloi;
  model dthtime*dthstat(0)=mlogrna cd4grp1 cd4grp2 combther decrs8 incrs8
        / risklimits;
  cd4level: test cd4grp1, cd4grp2;
  wk8resp: test decrs8, incrs8;
```

Notes:

- The "risklimits" option on the model statement provides 95% confidence intervals using Method II from page 2. (i.e., forming a CI for the log HR (beta), and then exponentiating the bounds)
- The "test" statement has the following form:

  `Label: test varname1, varname2, ..., varnamek;`

  for a k df Wald chi-square test of whether the k coefficients are all equal to 0.
- We can use the same approach described by Freedman to assess the effects of intermediate endpoints (incrs8, decrs8) on the treatment effect (i.e., assess their use as surrogate markers). The percentage of treatment effect explained, γ, is estimated by:

$$\hat{\gamma} \;=\; 1 - \frac{\hat{\beta}_{trt,M2}}{\hat{\beta}_{trt,M1}}$$

  where M1 is the model without the intermediate endpoint and M2 is the model with the marker.

OUTPUT FROM PROC PHREG (Model 1)

Proportional hazards regression model for time to Death
Baseline viral load and CD4 predictors

Data Set: WORK.ALLOI
Dependent Variable: DTHTIME Time to death (days)
Censoring Variable: DTHSTAT Death status (1=died,0=censored)
Censoring Value(s): 0
Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

Total	Event	Censored	Percent Censored
690	89	601	87.10

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	1072.543	924.167	148.376 with 4 DF (p=0.0001)
Score	.	.	189.702 with 4 DF (p=0.0001)
Wald	.	.	127.844 with 4 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
MLOGRNA	1	0.833237	0.17808	21.89295	0.0001
CD4GRP1	1	2.364612	0.32436	53.14442	0.0001
CD4GRP2	1	1.171137	0.34434	11.56739	0.0007
COMBTHER	1	-0.497161	0.24389	4.15520	0.0415

OUTPUT FROM PROC PHREG, continued

Output from "risklimits" and "test" statements

Analysis of Maximum Likelihood Estimates

Conditional Risk Ratio and 95% Confidence Limits

Variable	Risk Ratio	Lower	Upper	Label
MLOGRNA	2.301	1.623	3.262	log baseline rna (roche assay)
CD4GRP1	10.640	5.634	20.093	CD4<=100
CD4GRP2	3.226	1.643	6.335	100<CD4<=200
COMBTHER	0.608	0.377	0.981	Combination therapy with AZT/ddI/ddC/Nvp

Linear Hypotheses Testing

Label	Wald Chi-Square	DF	Pr > Chi-Square
CD4LEVEL	55.0794	2	0.0001

OUTPUT FROM PROC PHREG, (Model 2)

Proportional hazards regression model for time to Death
Baseline viral load and CD4 predictors

Data Set: WORK.ALLOI
Dependent Variable: DTHTIME Time to death (days)
Censoring Variable: DTHSTAT Death status (1=died,0=censored)
Censoring Value(s): 0
Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

Total	Event	Censored	Percent Censored
690	89	601	87.10

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	1072.543	912.009	160.535 with 6 DF (p=0.0001)
Score	.	.	198.537 with 6 DF (p=0.0001)
Wald	.	.	132.091 with 6 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square
MLOGRNA	1	0.893838	0.18062	24.48880	0.0001
CD4GRP1	1	2.023005	0.33594	36.26461	0.0001
CD4GRP2	1	1.001046	0.34907	8.22394	0.0041
COMBTHER	1	-0.456506	0.24687	3.41950	0.0644
DECRS8	1	-0.410919	0.26383	2.42579	0.1194
INCRS8	1	-0.834101	0.32884	6.43367	0.0112

OUTPUT FROM PROC PHREG, continued

Output from "risklimits" and "test" statements

Analysis of Maximum Likelihood Estimates

Conditional Risk Ratio and
95% Confidence Limits

Variable	Risk Ratio	Lower	Upper	Label
MLOGRNA	2.444	1.716	3.483	log baseline rna (roche assay)
CD4GRP1	7.561	3.914	14.606	CD4<=100
CD4GRP2	2.721	1.373	5.394	100<CD4<=200
COMBTHER	0.633	0.390	1.028	Combination therapy with AZT/ddI/ddC/Nvp
DECRS8	0.663	0.395	1.112	Decrease>=0.5 log rna at week 8?
INCRS8	0.434	0.228	0.827	Increase>=50 CD4 cells, week 8?

Linear Hypotheses Testing

Label	Wald Chi-Square	DF	Pr > Chi-Square
CD4LEVEL	37.6833	2	0.0001
WK8RESP	10.4312	2	0.0054

The percentage of treatment effect explained by including the RNA and CD4 response to treatment by Week 8 is:

$$\hat{\gamma} \;=\; 1 - \frac{-0.456}{-0.497} \approx 0.08$$

or 8%. The percentage of treatment effect on time to first opportunistic infection or death is much higher (about 24%).

II. Predicted Survival using PH

The Cox PH model says that $\lambda_i(t, \mathbf{Z}) = \lambda_0(t) \exp(\boldsymbol{\beta}\mathbf{Z})$. What does this imply about the survival function, $S_z(t)$, for the i-th individual with covariates $\mathbf{Z}_i$?

For the baseline (reference) group, we have:

$$S_0(t) = e^{-\int_0^t \lambda_0(u)du} = e^{-\Lambda_0(t)}$$

This is by definition of a survival function (see intro notes).

For the i-th patient with covariates $\mathbf{Z}_i$, we have:

$$\begin{aligned} S_i(t) &= e^{-\int_0^t \lambda_i(u)du} = e^{-\Lambda_i(t)} \\ &= e^{-\int_0^t \lambda_0(u) \exp(\boldsymbol{\beta}\mathbf{Z}_i)du} \\ &= e^{-\exp(\boldsymbol{\beta}\mathbf{Z}_i) \int_0^t \lambda_0(u)du} \\ &= \left[e^{-\int_0^t \lambda_0(u)du}\right]^{\exp(\boldsymbol{\beta}\mathbf{Z}_i)} \\ &= \left[S_0(t)\right]^{\exp(\boldsymbol{\beta}\mathbf{Z}_i)} \end{aligned}$$

(This uses the mathematical relationship $[e^b]^a = e^{ab}$)

Say we are interested in the survival pattern for single males in the nursing home study. Based on the previous formula, if we had an estimate for the survival function in the reference group, i.e., $\hat{S}_0(t)$, we could get estimates of the survival function for any set of covariates $\mathbf{Z}_i$.

How can we estimate the survival function, $S_0(t)$?

We could use the KM estimator, but there are a few disadvantages of that approach:

- It would only use the survival times for observations contained in the reference group, and not all the rest of the survival times.
- It would tend to be somewhat choppy, since it would reflect the smaller sample size of the reference group.
- It's possible that there are no subjects in the dataset who are in the "reference" group (ex. say covariates are age and sex; there is no one of age=0 in our dataset).

Instead, we will use a baseline hazard estimator which takes advantage of the proportional hazards assumption to get a smoother estimate.

$$\hat{S}_i(t) \;=\; [\hat{S}_0(t)]^{\exp(\widehat{\boldsymbol{\beta}}\mathbf{Z}_i)}$$

Using the above formula, we substitute $\widehat{\boldsymbol{\beta}}$ based on fitting the Cox PH model, and calculate $\hat{S}_0(t)$ by one of the following approaches:

- Breslow estimator (Stata)
- Kalbfleisch/Prentice estimator (SAS)

(1) **Breslow Estimator:**

$$\hat{S}_0(t) = \exp^{-\hat{\Lambda}_0(t)}$$

where $\hat{\Lambda}_0(t)$ is the estimated cumulative baseline hazard:

$$\hat{\Lambda}(t) = \sum_{j:\tau_j<t} \left(\frac{d_j}{\Sigma_{k\in\mathcal{R}(\tau_j)} \exp(\beta_1 Z_{1k} + \ldots \beta_p Z_{pk})} \right)$$

(2) **Kalbfleisch/Prentice Estimator**

$$\hat{S}_0(t) = \prod_{j:\tau_j<t} \hat{\alpha}_j$$

where $\hat{\alpha}_j, j = 1, ...d$ are the MLE's obtained by assuming that $S(t; Z)$ satisfies

$$S(t; Z) = [S_0(t)]^{e^{\beta Z}} = \left[\prod_{j:\tau_j<t} \alpha_j \right]^{e^{\beta Z}} = \prod_{j:\tau_j<t} \alpha_j^{e^{\beta Z}}$$

Breslow Estimator: further motivation

The Breslow estimator is based on extending the concept of the Nelson-Aalen estimator to the proportional hazards model.

Recall that for a single sample with no covariates, the **Nelson-Aalen Estimator** of the cumulative hazard is:

$$\hat{\Lambda}(t) = \sum_{j:\tau_j<t} \frac{d_j}{r_j}$$

where d_j and r_j are the number of deaths and the number at risk, respectively, at the j-th death time.

When there are covariates and assuming the PH model above, one can generalize this to estimate the cumulative baseline hazard by adjusting the denominator:

$$\hat{\Lambda}(t) = \sum_{j:\tau_j<t} \left(\frac{d_j}{\Sigma_{k\in\mathcal{R}(\tau_j)} \exp(\beta_1 Z_{1k} + \ldots \beta_p Z_{pk})} \right)$$

Heuristic: The expected number of failures in $(t, t+\delta t)$ is

$$d_j \approx \delta t \times \sum_{k\in\mathcal{R}(t)} \lambda_0(t) exp(z_k \hat{\beta})$$

Hence,

$$\delta t \times \lambda_0(t_j) \approx \frac{d_j}{\Sigma_{k\in\mathcal{R}(t)} exp(z_k \hat{\beta})}$$

Kalbfleisch/Prentice Estimator: further motivation

This method is analogous to the Kaplan-Meier Estimator. Consider a discrete time model with hazard $(1 - \alpha_j)$ at the j-th observed death time.

(Note: we use $\alpha_j = (1 - \lambda_j)$ to simplify the algebra!)

Thus, for someone with z=0, the survivorship function is

$$S_0(t) = \prod_{j:\tau_j<t} \alpha_j$$

and for someone with $Z \neq 0$, it is:

$$S(t; Z) = S_0(t)^{e^{\beta Z}} = \left[\prod_{j:\tau_j<t} \alpha_j\right]^{e^{\beta Z}} = \prod_{j:\tau_j<t} \alpha_j^{e^{\beta Z}}$$

The likelihood contributions under this model are:

- for someone censored at t: $S(t; Z)$
- for someone who fails at t_j:

$$S(t_{(j-1)}; Z) - S(t_j; Z) = \left[\prod_{k<j} \alpha_j\right]^{e^{\beta z}} [1 - \alpha_j^{e^{\beta Z}}]$$

The solution for α_j satisfies:

$$\sum_{k\in\mathcal{D}_j} \frac{\exp(Z_k\beta)}{1 - \alpha_j^{\exp(Z_k\beta)}} = \sum_{k\in\mathcal{R}_j} \exp(Z_k\beta)$$

(Note what happens when $Z = 0$)

Obtaining $\hat{S}_0(t)$ from software packages

- Stata provides the Breslow estimator of $S_0(t; Z)$, but not predicted survivals at specified covariate values..... you have to construct these yourself

- SAS uses the Kalbfleisch/Prentice estimator of the baseline hazard, and can provide estimates of survival at arbitrary values of the covariates with a little bit of programming.

In practice, they are **incredibly** close! (see Fleming and Harrington 1984, *Communications in Statistics*)

Using Stata to Predict Survival

The Stata command `basesurv` calculates the predicted survival values for the reference group, i.e., those subjects with all covariates=0.

(1) **Baseline Survival:**
To obtain the estimated baseline survival $\hat{S}_0(t)$, follow the example below (for the nursing home data):

```
. use nurshome

. stset los fail

. stcox married health, basesurv(prsurv)

. sort los

. list los prsurv
```

Estimating the Baseline Survival with Stata

```
            los      prsurv

   1.         1   .99252899
   2.         1   .99252899
   3.         1   .99252899
   4.         1   .99252899
   5.         1   .99252899
.
.
.
  22.         1   .99252899
  23.         2   .98671824
  24.         2   .98671824
  25.         2   .98671824
  26.         2   .98671824
  27.         2   .98671824
  28.         2   .98671824
  29.         2   .98671824
  30.         2   .98671824
  31.         2   .98671824
  32.         2   .98671824
  33.         2   .98671824
  34.         2   .98671824
  35.         2   .98671824
  36.         2   .98671824
  37.         2   .98671824
  38.         2   .98671824
  39.         2   .98671824
  40.         3   .98362595
  41.         3   .98362595
.
.
.
```

Stata creates a predicted baseline survival estimate for every observed event time in the dataset, even if there are duplicates.

(2) **Predicted Survival for Subgroups**

To obtain the estimated survival $\hat{S}_i(t)$ for any other subgroup (i.e., not the reference or baseline group), follow the Stata commands below:

```
. predict betaz, xb

. gen newterm=exp(betaz)

. gen predsurv=prsurv^newterm

. sort married health los

. list married health los predsurv
```

Predicting Survival for Subgroups with Stata

```
       married     health        los   predsurv

  1.         0          2          1   .9896138
  8.         0          2          2    .981557
 11.         0          2          3   .9772769
 13.         0          2          4   .9691724
 16.         0          2          5   .9586483
.....................................................................
300.         0          3          1   .9877566
302.         0          3          2   .9782748
304.         0          3          3   .9732435
305.         0          3          4   .9637272
312.         0          3          5   .9513916
.....................................................................
768.         0          4          1   .9855696
777.         0          4          2   .9744162
779.         0          4          3   .9685058
781.         0          4          4   .9573418
785.         0          4          5   .9428996
.
.
.
1468.        1          4          1   .9806339
1469.        1          4          2   .9657326
1472.        1          4          3   .9578599
1473.        1          4          5   .9239448
.....................................................................
1559.        1          5          1   .9771894
1560.        1          5          2   .9596928
1562.        1          5          3   .9504684
1564.        1          5          4   .9331349
```

Using SAS to Predict Survival

The SAS command BASELINE calculates the predicted survival values at the event times for a given set of covariate values.

(1) To get the estimated baseline survival $\hat{S}_0(t)$, create a dataset with 0's for values of all covariates in the model

(2) To get the estimated survival $\hat{S}_i(t)$ for any other subgroup (i.e., not the reference or baseline group), create a data set which inputs the baseline values of the covariates for the subgroup of interest.

For either case, we then supply the corresponding dataset name to the BASELINE command under PROC PHREG.

By giving the input dataset several lines, each corresponding to a different combination of covariate values, we can compute predicted survival values for more than one group at once.

(1) Baseline Survival Estimate

(note that the baseline survival function does not correspond to any observations in our sample, since health status values range from 2-5)

```
*** Estimating Baseline Survival Function under PH;
data inrisks;
  input married health;
  cards;
0 0
;

proc phreg data=pop out=survres;
  model los*fail(0)=married health;
  baseline covariates=inrisks out=outph survival=ps/nomean;

proc print data=outph;
title1 'Nursinghome data: Baseline Survival Estimate';
```

Estimating the Baseline Survival with SAS

Nursinghome data: Baseline Survival Estimate

OBS	MARRIED	HEALTH	LOS	PS
1	0	0	0	1.00000
2	0	0	1	0.99253
3	0	0	2	0.98672
4	0	0	3	0.98363
5	0	0	4	0.97776
6	0	0	5	0.97012
7	0	0	6	0.96488
8	0	0	7	0.95856
9	0	0	8	0.95361
10	0	0	9	0.94793
11	0	0	10	0.94365
12	0	0	11	0.93792
13	0	0	12	0.93323
14	0	0	13	0.92706
15	0	0	14	0.92049
16	0	0	15	0.91461
17	0	0	16	0.91017
18	0	0	17	0.90534
19	0	0	18	0.90048
20	0	0	19	0.89635
21	0	0	20	0.89220
22	0	0	21	0.88727
23	0	0	22	0.88270

.
.
.

(2) Predicted Survival Estimate for Subgroup

The following SAS commands will generate the predicted survival probability for each combination of covariates, at every observed event time in the dataset.

```
*** Estimating Baseline Survival Function under PH;
data inrisks;
  input married health;
  cards;
0 2
0 5
1 2
1 5
;

proc phreg data=pop out=survres;
  model los*fail(0)=married health;
  baseline covariates=inrisks out=outph survival=ps/nomean;

proc print data=outph;
title1 'Nursinghome data: predicted survival by subgroup';
```

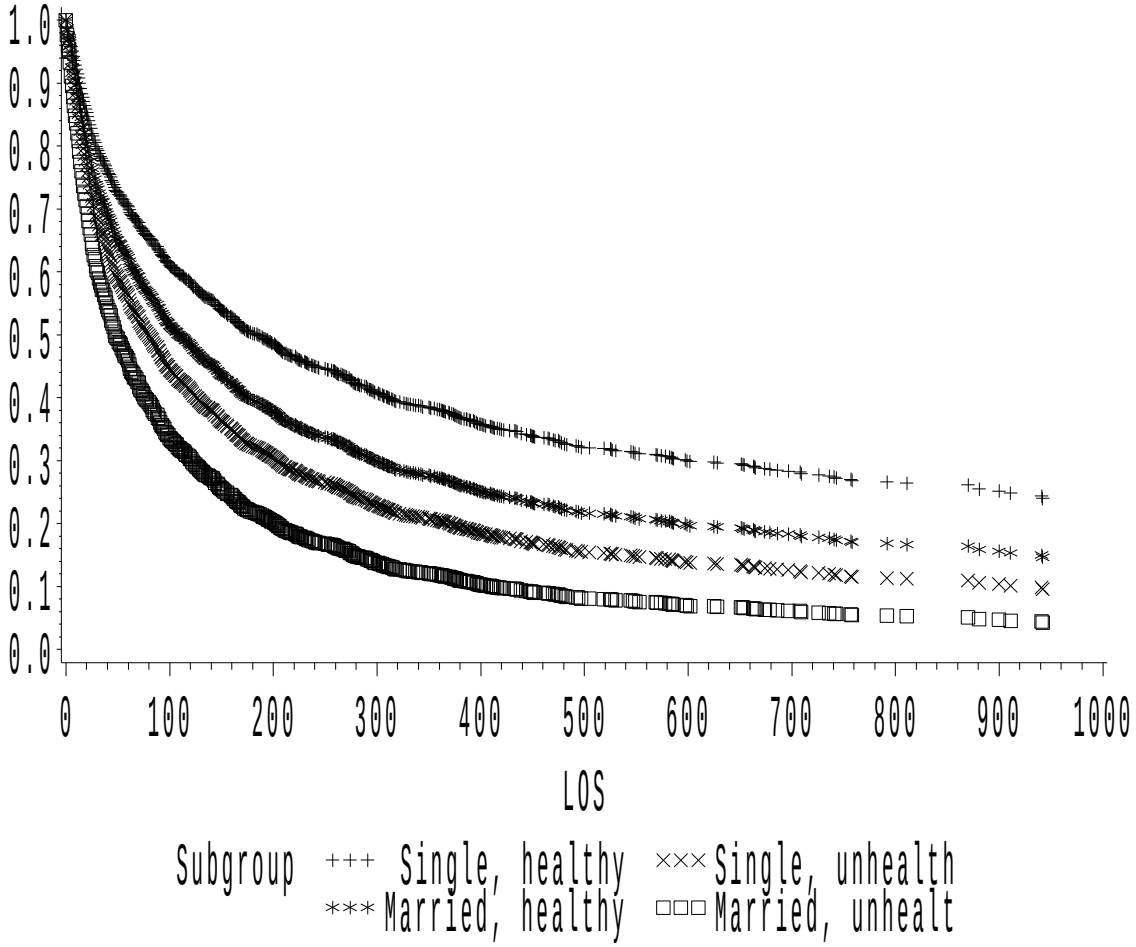
Survival Estimates by Marital and Health Status

Nursinghome data: Predicted Survival by Subgroup

OBS	MARRIED	HEALTH	LOS	PS
1	0	2	0	1.00000
2	0	2	1	0.98961
3	0	2	2	0.98156
4	0	2	3	0.97728
...	...	...	...	...
171	0	2	184	0.50104
172	0	2	185	0.49984
...	...	...	...	...
396	0	5	0	1.00000
397	0	5	1	0.98300
398	0	5	2	0.96988
399	0	5	3	0.96295
...	...	...	...	...
474	0	5	78	0.50268
475	0	5	80	0.49991
...	...	...	...	...
791	1	2	0	1.00000
792	1	2	1	0.98605
793	1	2	2	0.97527
794	1	2	3	0.96955
...	...	...	...	...
897	1	2	108	0.50114
898	1	2	109	0.49986
...	...	...	...	...
1186	1	5	0	1.00000
1187	1	5	1	0.97719
1188	1	5	2	0.95969
1189	1	5	3	0.95047
...	...	...	...	...
1233	1	5	47	0.50519
1234	1	5	48	0.49875

We can get a visual picture of what the proportional hazards assumption implies by looking at these four subgroups

Nursinghome data: Predicted Survival by Subgroup



III. Predicted medians and P-year survival

Predicted Medians

Suppose we want to find the predicted median survival for an individual with a specified combination of covariates (e.g., a single person with health status 5).

Three possible approaches:

(1) Calculate the median from the subset of individuals with the specified covariate combination (using KM approach)

(2) Generate predicted survival curves for each combination of covariates, and obtain the medians directly

OBS	MARRIED	HEALTH	LOS	PREDSURV
171	0	2	184	0.50104
172	0	2	185	0.49984
474	0	5	78	0.50268
475	0	5	80	0.49991
897	1	2	108	0.50114
898	1	2	109	0.49986
1233	1	5	47	0.50519
1234	1	5	48	0.49875

Recall that previously we defined the median as the *smallest* value of t for which $\hat{S}(t) \leq 0.5$, so the medians from above would be 185, 80, 109, and 48 days for single healthy, single unhealthy, married healthy, and married unhealthy, respectively.

(3) Generate the predicted survival curve from the estimated baseline hazard, as follows:

We want the estimated median (M) for an individual with covariates $\mathbf{Z}_i$. We know

$$S(M; Z) = [S_0(M)]^{e^{\beta Z_i}} = 0.5$$

Hence, M satisfies (multiplying both sides by $e^{-\beta Z_i}$):

$$S_0(M) = [0.5]^{e^{-\beta Z}}$$

Ex. Suppose we want to estimate the median survival for a single unhealthy subject from the nursing home data. The reciprocal of the hazard ratio for unhealthy (health=5) is: $e^{-0.165*5} = 0.4373$, (where $\hat{\beta} = 0.165$ for health status)

So, we want M such that $S_0(M) = (0.5)^{0.4373} = 0.7385$

So the median for single unhealthy subject is the 73.8^{th} percentile of the baseline group.

OBS	MARRIED	HEALTH	LOS	PREDSURV
79	0	0	78	0.74028
80	0	0	80	0.73849
81	0	0	81	0.73670

So the estimated median would still be 80 days. Note: similar logic can be followed to estimate other quantiles besides the median.

Estimating P-year survival

Suppose we want to find the P-year survival rate for an individual with a specified combination of covariates, $\hat{S}(P; \mathbf{Z}_i)$

For an individual with $\mathbf{Z}_i = 0$, the P-year survival can be obtained from the baseline survivorship function, $\hat{S}_0(P)$

For individuals with $\mathbf{Z}_i \neq 0$, it can be obtained as:

$$\hat{S}(P; \mathbf{Z}_i) = [\hat{S}_0(P)]^{e^{\widehat{\boldsymbol{\beta}}\mathbf{Z}_i}}$$

Notes:

- Although I say "P-year" survival, the units of time in a particular dataset may be days, weeks, or months. The answer here will be in the same units of time as the original data.
- If $\widehat{\boldsymbol{\beta}}\mathbf{Z}_i$ is positive, then the P-year survival rate for the i-th individual will be lower than for a baseline individual.

  Why is this true?

Model Selection in Survival Analysis

Suppose we have a censored survival time that we want to model as a function of a (possibly large) set of covariates. Two important questions are:

- How to decide which covariates to use
- How to decide if the final model fits well

To address these topics, we'll consider a new example:

Survival of Atlantic Halibut - Smith et al

Obs #	*Surv*ival *Time* (min)	*Censor*ing Indicator	*Tow Dur*ation (min.)	Diff in *Depth*	*Length* of Fish (cm)	*Handling* Time (min.)	Total *log(catch)* ln(weight)
100	353.0	1	30	15	39	5	5.685
109	111.0	1	100	5	44	29	8.690
113	64.0	0	100	10	53	4	5.323
116	500.0	1	100	10	44	4	5.323
⋮							

Hosmer & Lemeshow
Chapter 5: Model Development
Chapter 6: Assessment of Model Adequacy
(sections 6.1-6.2)

Process of Model Selection

Collett (Section 3.6) has an excellent discussion of various approaches for model selection. In practice, model selection proceeds through a combination of

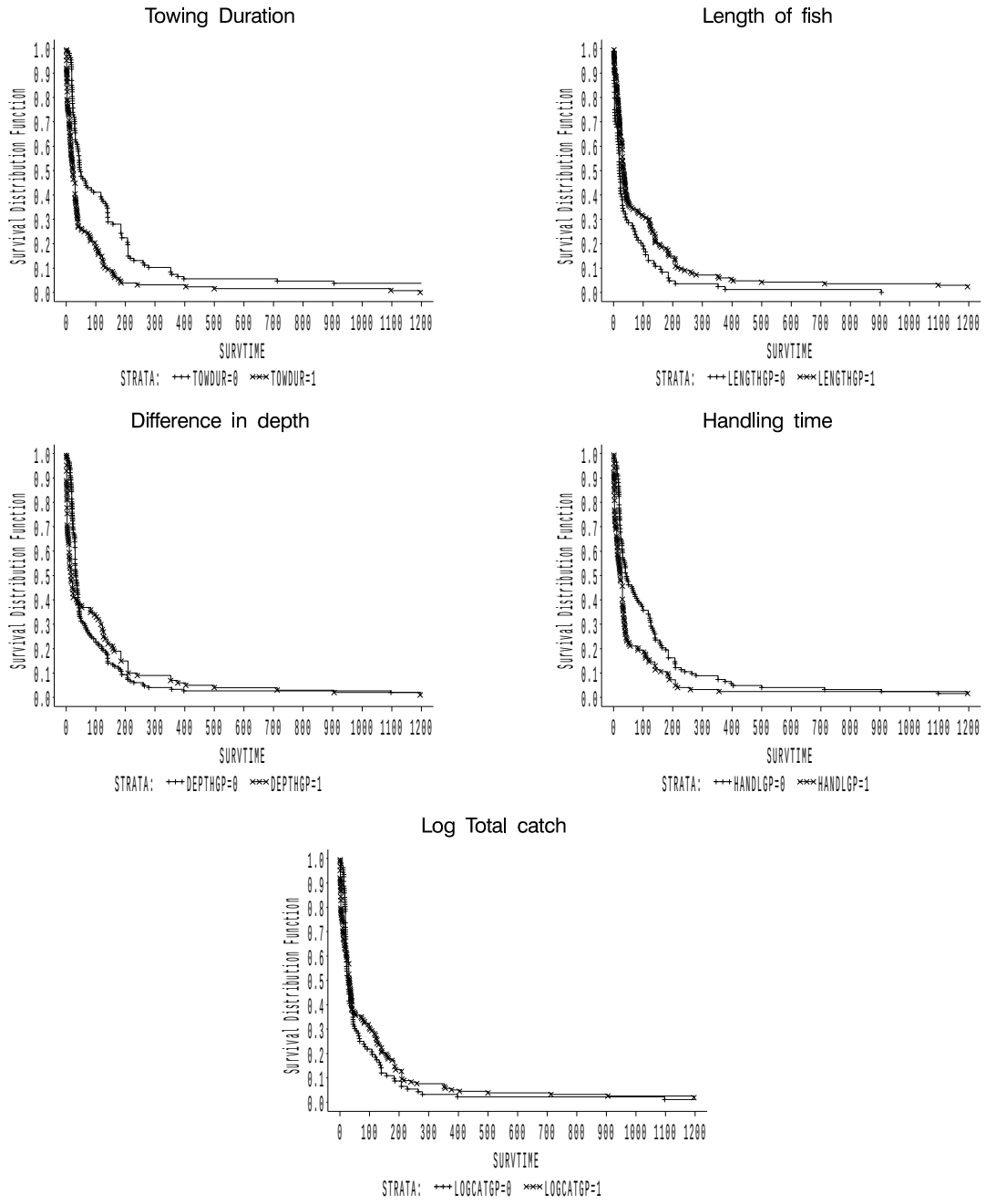
- knowledge of the science
- trial and error, common sense
- automatic variable selection procedures
  - forward selection
  - backward selection
  - stepwise selection

Many advocate the approach of first doing a univariate analysis to "screen" out potentially significant variables for consideration in the multivariate model (see Collett).

Let's start with this approach.

Univariate KM plots of Atlantic Halibut survival

(continuous variables have been dichotomized)



Which covariates look like they might be important?

Automatic Variable selection procedures in Stata and SAS

Statistical Software:

- Stata: `sw` command before `cox` command
- SAS: `selection=` option on model statement of `proc phreg`

Options:

(1) forward

(2) backward

(3) stepwise

(4) best subset (SAS only, using `score` option)

One drawback of these options is that they can only handle variables one at a time. When might that be a disadvantage?

Collett's Model Selection Approach
Section 3.6.1

This approach assumes that all variables are considered to be on an equal footing, and there is no *a priori* reason to include any specific variables (like treatment).

Approach:

(1) Fit a univariate model for each covariate, and identify the predictors significant at some level p_1, say 0.20.

(2) Fit a multivariate model with all significant univariate predictors, and use *backward* selection to eliminate non-significant variables at some level p_2, say 0.10.

(3) Starting with final step (2) model, consider each of the non-significant variables from step (1) using *forward* selection, with significance level p_3, say 0.10.

(4) Do final pruning of main-effects model (omit variables that are non-significant, add any that are significant), using *stepwise* regression with significance level p_4. At this stage, you may also consider adding interactions between any of the main effects currently in the model, under the hierarchical principle.

Collett recommends using a likelihood ratio test for all variable inclusion/exclusion decisions.

Stata Command for Forward Selection:

Forward Selection $\Longrightarrow$ use $pe(\alpha)$ option, where α is the significance level for entering a variable into the model.

```
. use halibut

. stset survtime censor

. sw cox survtime towdur depth length handling logcatch,
> dead(censor)  pe(.05)

                      begin with empty model

p = 0.0000 <  0.0500  adding   handling
p = 0.0000 <  0.0500  adding   logcatch
p = 0.0010 <  0.0500  adding   towdur
p = 0.0003 <  0.0500  adding   length

Cox Regression -- entry time 0                        Number of obs =    294
                                                      chi2(4)       =  84.14
                                                      Prob > chi2   = 0.0000
Log Likelihood = -1257.6548                           Pseudo R2     = 0.0324

------------------------------------------------------------------------------
survtime |
  censor |      Coef.   Std. Err.       z     P>|z|     [95% Conf. Interval]
---------+--------------------------------------------------------------------
handling |   .0548994   .0098804      5.556   0.000     .0355341    .0742647
logcatch |  -.1846548    .051015     -3.620   0.000     .2846423   -.0846674
  towdur |   .5417745   .1414018      3.831   0.000     .2646321     .818917
  length |  -.0366503   .0100321     -3.653   0.000    -.0563129   -.0169877
------------------------------------------------------------------------------
```

Stata Command for Backward Selection:

Backward Selection $\Longrightarrow$ use $pr(\alpha)$ option, where α is the significance level for a variable to remain in the model.

```
. sw cox survtime towdur depth length handling logcatch,
> dead(censor) pr(.05)


                      begin with full model

p = 0.1991 >= 0.0500  removing depth

Cox Regression -- entry time 0                    Number of obs =    294
                                                  chi2(4)       =  84.14
                                                  Prob > chi2   = 0.0000
Log Likelihood = -1257.6548                       Pseudo R2     = 0.0324

------------------------------------------------------------------------------
survtime |
  censor |      Coef.   Std. Err.       z     P>|z|   [95% Conf. Interval]
---------+--------------------------------------------------------------------
  towdur |   .5417745   .1414018      3.831   0.000   .2646321     .818917
logcatch |  -.1846548    .051015     -3.620   0.000  -.2846423   -.0846674
  length |  -.0366503   .0100321     -3.653   0.000  -.0563129   -.0169877
handling |   .0548994   .0098804      5.556   0.000   .0355341    .0742647
------------------------------------------------------------------------------
```

Stata Command for Stepwise Selection:

Stepwise Selection $\Longrightarrow$ use both $pe(.)$ and $pr(.)$ options, with $pr(.) > pe(.)$

```
. sw cox survtime towdur depth length handling logcatch,
> dead(censor) pr(0.10) pe(0.05)


                     begin with full model

p = 0.1991 >= 0.1000  removing depth

Cox Regression -- entry time 0                    Number of obs =    294
                                                  chi2(4)       =  84.14
                                                  Prob > chi2   = 0.0000
Log Likelihood = -1257.6548                       Pseudo R2     = 0.0324

------------------------------------------------------------------------------
survtime |
  censor |      Coef.   Std. Err.       z     P>|z|  [95% Conf. Interval]
---------+--------------------------------------------------------------------
  towdur |   .5417745   .1414018      3.831   0.000  .2646321     .818917
handling |   .0548994   .0098804      5.556   0.000  .0355341    .0742647
  length |  -.0366503   .0100321     -3.653   0.000 -.0563129   -.0169877
logcatch |  -.1846548    .051015     -3.620   0.000 -.2846423   -.0846674
------------------------------------------------------------------------------
```

It is also possible to do forward stepwise regression by including both $pr(.)$ and $pe(.)$ options with `forward` option

SAS programming statements for model selection

```
data fish;
 infile 'fish.dat';
 input ID SURVTIME CENSOR TOWDUR DEPTH LENGTH HANDLING LOGCATCH;
run;

title 'Survival of Atlantic Halibut';
*** automatic variable selection procedures;
proc phreg data=fish;
  model survtime*censor(0)= towdur depth length handling logcatch
        /selection=stepwise slentry=0.1 slstay=0.1 details;
  title2 'Stepwise selection';
run;

proc phreg data=fish;
  model survtime*censor(0)= towdur depth length handling logcatch
        /selection=forward slentry=0.1 details;
  title2 'Forward selection';
run;

proc phreg data=fish;
  model survtime*censor(0)= towdur depth length handling logcatch
        /selection=backward slstay=0.1 details;
  title2 'Backward selection';
run;

proc phreg data=fish;
  model survtime*censor(0)= towdur depth length handling logcatch
        /selection=score;
  title2 'Best subsets selection';
run;
```

Final model for stepwise selection approach

Survival of Atlantic Halibut
Stepwise selection

The PHREG Procedure

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
TOWDUR	1	0.007740	0.00202	14.68004	0.0001	1.008
LENGTH	1	-0.036650	0.01003	13.34660	0.0003	0.964
HANDLING	1	0.054899	0.00988	30.87336	0.0001	1.056
LOGCATCH	1	-0.184655	0.05101	13.10166	0.0003	0.831

Analysis of Variables Not in the Model

Variable	Score Chi-Square	Pr > Chi-Square
DEPTH	1.6661	0.1968

Residual Chi-square = 1.6661 with 1 DF (p=0.1968)

NOTE: No (additional) variables met the 0.1 level for entry into the model.

Summary of Stepwise Procedure

Step	Variable Entered	Variable Removed	Number In	Score Chi-Square	Wald Chi-Square	Pr > Chi-Square
1	HANDLING		1	47.1417	.	0.0001
2	LOGCATCH		2	18.4259	.	0.0001
3	TOWDUR		3	11.0191	.	0.0009
4	LENGTH		4	13.4222	.	0.0002

Output from PROC SAS "score" option

```
  NUMBER OF        SCORE   VARIABLES INCLUDED
  VARIABLES        VALUE   IN MODEL

          1      47.1417   HANDLING
          1      29.9604   TOWDUR
          1      12.0058   LENGTH
          1       4.2185   DEPTH
          1       1.4795   LOGCATCH
--------------------------------
          2      65.6797   HANDLING LOGCATCH
          2      59.9515   TOWDUR HANDLING
          2      56.1825   LENGTH HANDLING
          2      51.6736   TOWDUR LENGTH
          2      47.2229   DEPTH HANDLING
          2      32.2509   TOWDUR LOGCATCH
          2      30.6815   TOWDUR DEPTH
          2      16.9342   DEPTH LENGTH
          2      14.4412   LENGTH LOGCATCH
          2       9.1575   DEPTH LOGCATCH
-------------------------------------
          3      76.8829   LENGTH HANDLING LOGCATCH
          3      76.3454   TOWDUR HANDLING LOGCATCH
          3      75.5291   TOWDUR LENGTH HANDLING
          3      69.0334   DEPTH HANDLING LOGCATCH
          3      60.0340   TOWDUR DEPTH HANDLING
          3      56.4207   DEPTH LENGTH HANDLING
          3      55.8374   TOWDUR LENGTH LOGCATCH
          3      52.4130   TOWDUR DEPTH LENGTH
          3      34.7563   TOWDUR DEPTH LOGCATCH
          3      24.2039   DEPTH LENGTH LOGCATCH
-----------------------------------------
          4      94.0062   TOWDUR LENGTH HANDLING LOGCATCH
          4      81.6045   DEPTH LENGTH HANDLING LOGCATCH
          4      77.8234   TOWDUR DEPTH HANDLING LOGCATCH
          4      75.5556   TOWDUR DEPTH LENGTH HANDLING
          4      59.1932   TOWDUR DEPTH LENGTH LOGCATCH
-------------------------------------------------
          5      96.1287   TOWDUR DEPTH LENGTH HANDLING LOGCATCH
-------------------------------------------------------
```

Best multivariate model for all 3 options

Survival of Atlantic Halibut
Best Multivariate Model

The PHREG Procedure

Data Set: WORK.FISH
Dependent Variable: TIME
Censoring Variable: CENSOR
Censoring Value(s): 0
Ties Handling: BRESLOW

Summary of the Number of
Event and Censored Values

Total	Event	Censored	Percent Censored
294	273	21	7.14

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	2599.449	2515.310	84.140 with 4 DF (p=0.0001)
Score	.	.	94.006 with 4 DF (p=0.0001)
Wald	.	.	90.247 with 4 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
TOWDUR	1	0.007740	0.00202	14.68004	0.0001	1.008
LENGTH	1	-0.036650	0.01003	13.34660	0.0003	0.964
HANDLING	1	0.054899	0.00988	30.87336	0.0001	1.056
LOGCATCH	1	-0.184655	0.05101	13.10166	0.0003	0.831

Notes:

- When the halibut data was analyzed with the forward, backward and stepwise options, the same final model was reached. However, this will not always be the case.

- Variables can be forced into the model using the `lockterm` option in Stata and the `include` option in SAS. Any variables that you want to force inclusion of must be listed first in your model statement.

- Stata uses the Wald test for both forward and backward selection, although it has an option to use the likelihood ratio test instead (`lrtest`). SAS uses the score test to decide what variables to add and the Wald test for what variables to remove.

- If you fit a range of models manually, you can apply the AIC criteria described by Collett:

  $$\text{minimize AIC} = -2 \log(\hat{L}) + (\alpha * q)$$

  where q is the number of unknown parameters in the model and α is typically between 2 and 6 (they suggest $\alpha = 3$).

  The model is then chosen which minimizes the AIC (similar to maximizing log-likelihood, but with a penalty for number of variables in the model)

Questions:

- When might we want to force certain variables into the model?

  (1) to examine interactions

  (2) to keep main effects in the model

  (3) to calculate a score test for a paricular effect

- Would it be possible to get different final models from SAS and Stata?

- Based on what we've seen in the behavior of Wald tests, would SAS or Stata be more likely to add a covariate to a model in a forward selection model?

- If we use the AIC criteria with $\alpha = 3$, how does that compare to the likelihood ratio test?

Assessing overall model fit

How do we know if the model fits well?

- Always look at univariate plots (Kaplan-Meiers)

  Construct a Kaplan-Meier survival plot for each of the important predictors, like the ones shown at the beginning of these notes.

- Check proportionality assumption (this will be the topic of the next lecture)

- **Check residuals!**

  (a) generalized (Cox-Snell)
  (b) martingale
  (c) deviance
  (d) Schoenfeld
  (e) weighted Schoenfeld

Residuals for survival data are slightly different than for other types of models, due to the censoring. Before we start talking about residuals, we need an important basic result:

Inverse CDF:

If T_i (the survival time for the i-th individual) has survivorship function $S_i(t)$, then the transformed random variable $S_i(T_i)$ (i.e., the survival function evaluated at the actual survival time T_i) should be from a uniform distribution on $[0, 1]$, and hence $-\log[S_i(T_i)]$ should be from a unit exponential distribution

More mathematically:

$$\begin{aligned} \text{If} \quad T_i &\sim S_i(t) \\ \text{then} \quad S_i(T_i) &\sim Uniform[0,1] \\ \text{and} \quad -\log S_i(T_i) &\sim Exponential(1) \end{aligned}$$

(a) Generalized (Cox-Snell) Residuals:

The implication of the last result is that if the model is correct, the estimated cumulative hazard for each individual at the time of their death or censoring should be like a censored sample from a unit exponential. This quantity is called the *generalized* or *Cox-Snell* residual.

Here is how the generalized residual might be used. Suppose we fit a PH model:

$$S(t;Z) = [S_0(t)]^{\exp(\beta Z)}$$

or, in terms of hazards:

$$\begin{aligned} \lambda(t;Z) &= \lambda_0(t)\exp(\beta Z) \\ &= \lambda_0(t)\exp(\beta_1 Z_1 + \beta_2 Z_2 + \cdots + \beta_k Z_k) \end{aligned}$$

After fitting, we have:

- $\hat{\beta}_1, \ldots, \hat{\beta}_k$
- $\hat{S}_0(t)$

So, for each person with covariates $\mathbf{Z}_i$, we can get

$$\hat{S}(t; \mathbf{Z}_i) = [\hat{S}_0(t)]^{\exp(\boldsymbol{\beta}\mathbf{Z}_i)}$$

This gives a predicted survival probability at each time t in the dataset (see notes from the previous lecture).

Then we can calculate

$$\hat{\Lambda}_i = -\log[\hat{S}(T_i; Z_i)]$$

In other words, first we find the predicted survival probability at the actual survival time for an individual, then log-transform it.

Example: Nursing home data

Say we have

- a single male
- with actual duration of stay of 941 days ($X_i = 941$)

We compute the entire distribution of survival probabilities for single males, and obtain $\hat{S}(941) = 0.260$.

$$-\log[\hat{S}(941, \text{single male})] \;=\; -\log(0.260) = 1.347$$

We repeat this for everyone in our dataset. These should be like a censored sample from an exponential (1) distribution if the model fits the data well.

Based on the properties of a unit exponential model

- plotting $-\log(\hat{S}(t))$ vs t should yield a straight line
- plotting $\log[-\log S(t)]$ vs $\log(t)$ should yield a straight line through the origin with slope=1.

To convince yourself of this, start with $S(t) = e^{-\lambda t}$ and calculate $\log[-\log S(t)]$. What do you get for the slope and intercept?

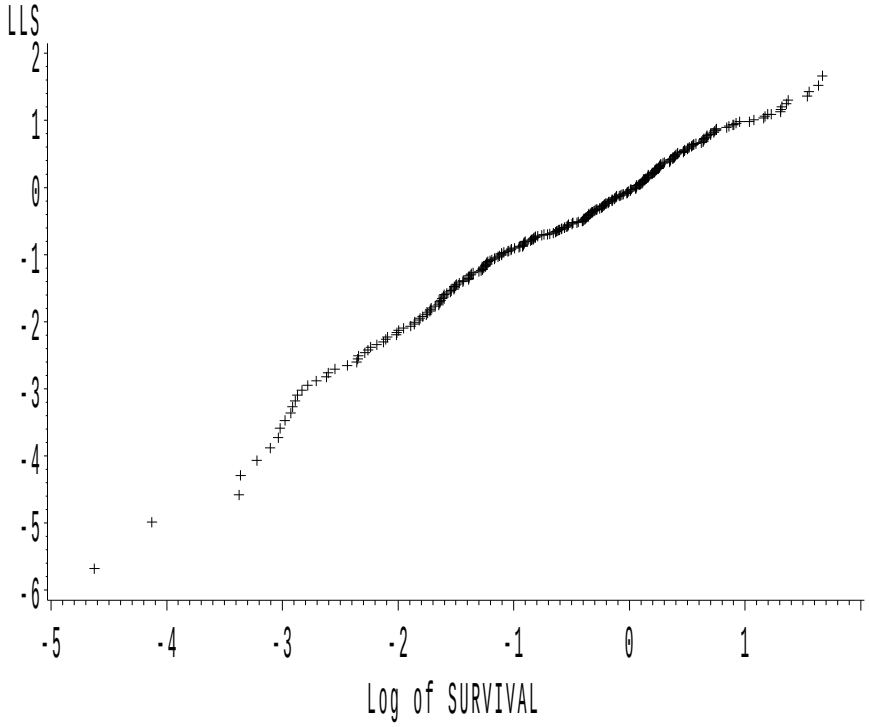
(Note: this does not necessarily mean that the underlying distribution of the original survival times is exponential!)

Obtaining the generalized residuals from Stata

- Fit a Cox PH model with the `stcox` command, along with the `mgale(`*`newvar`*`)` option
- Use the `predict` command with the `csnell` option
- Define a survival dataset using the Cox-Snell residuals as the "pseudo" failure times
- Calculate the estimated KM survival
- Take the $\log[-\log(S(t))]$ based on the above
- Generate the log of the Cox-Snell residuals
- Graph $\log[-\log S(t)]$ vs $\log(t)$

```
. stcox towdur handling length logcatch, mgale(mg)

. predict csres, csnell

. stset csres censor

. sts list

. sts gen survcs=s

. gen lls=log(-log(survcs))

. gen loggenr=log(csres)

. graph lls loggenr
```

Does the exponential model fit?



Allison states "Cox-Snell residuals... are not very informative for Cox models estimated by partial likelihood." He instead prefers deviance residuals (later).

Obtaining the generalized residuals from SAS

The **generalized residuals** can be obtained from SAS after fitting a PH model using the output statement with the **logsurv** option.

```
proc phreg data=fish;
  model survtime*censor(0) = towdur handling logcatch length;
  output out=phres logsurv=genres;

*** take negative log Pr(survival) at each persons survtime;
data phres;
  set phres;
  genres=-genres;

*** Now we treat the generalized residuals as the input dataset;
*** to evaluate whether the assumption of an exponential;
*** distribution is appropriate;
proc lifetest data=phres outsurv=survres;
  time genres*censor(0);

data survres;
  set survres;
  lls=log(-log(survival));
  loggenr=log(genres);

proc gplot data=survres;
  plot lls*loggenr;
run;
```

(b) Martingale Residuals

(see Fleming and Harrington, p.164)

Martingale residuals are defined for the i-th individual as:

$$r_i = \delta_i - \hat{\Lambda}(T_i)$$

Properties:

- r_i's have mean 0
- range of r_i's is between $-\infty$ and 1
- approximately uncorrelated (in large samples)
- **Interpretation:** - the residual r_i can be viewed as the difference between the observed number of deaths (0 or 1) for subject i between time 0 and T_i, and the expected numbers based on the fitted model.

The **martingale residuals** can be obtained from Stata using the `mgale` option shown previously.

Once the martingale residual is created, you can plot it versus the predicted log HR (i.e., $\boldsymbol{\beta}\mathbf{Z}_i$), or any of the individual covariates.

```
. stcox towdur handling length logcatch, mgale(mg)

. predict betaz=xb

. graph mg betaz

. graph mg logcatch

. graph mg towdur

. graph mg handling

. graph mg length
```

The **martingale residuals** can be obtained from SAS after fitting a PH model using the output statement with the **resmart** option.

Once you have them, you can

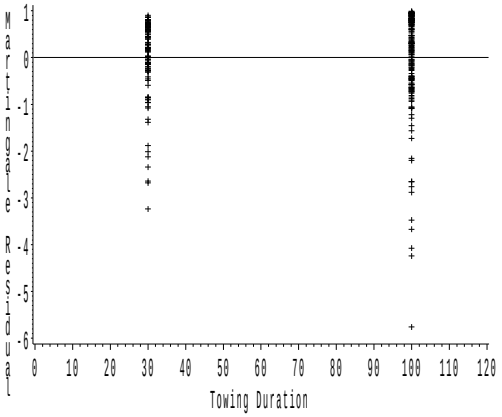
- plot against predicted values
- plot against covariates

```
proc phreg data=fish;
  model survtime*censor(0) = towdur handling logcatch length;
  output out=phres resmart=mres xbeta=xb;

proc gplot data=phres;
  plot mres*xb;              /* predicted values */
  plot mres*towdur;
  plot mres*handling;
  plot mres*logcatch;
  plot mres*length;
run;
```

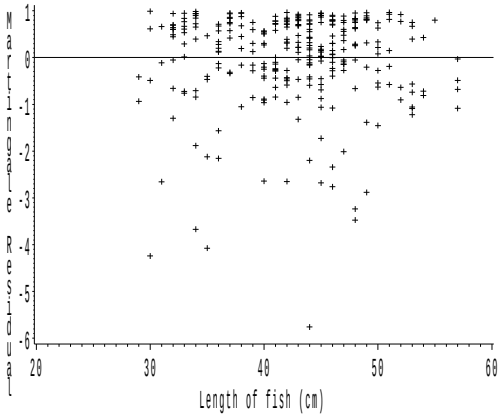
Allison still prefers the deviance residuals (next)

Martingale Residuals

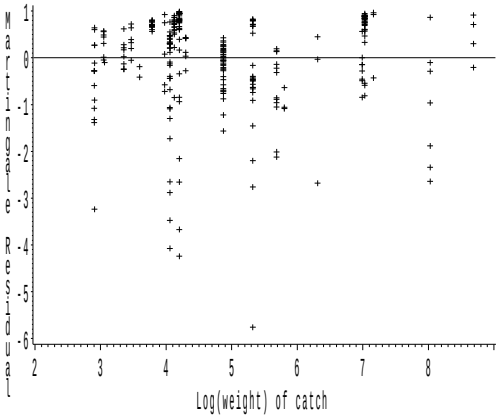
Martingale residuals vs towing duration



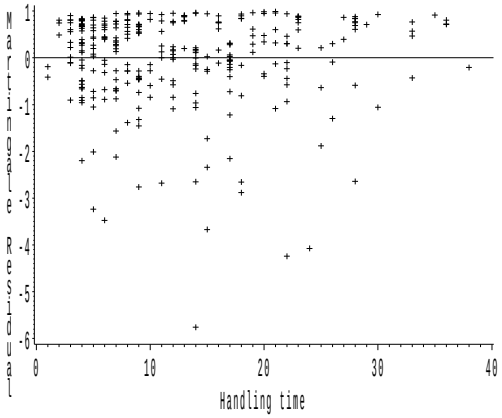
Martingale residuals vs length of fish



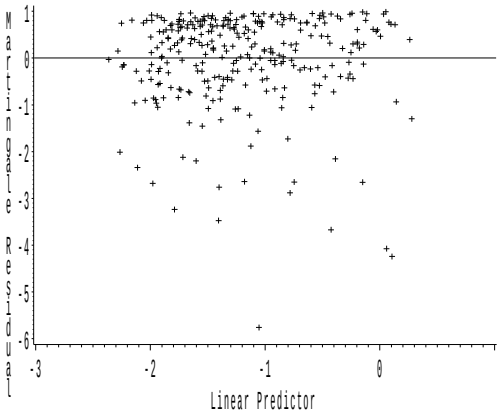
Martingale residuals vs log(catch)



Martingale residuals vs handling



Martingale residuals vs predicted values



(c) Deviance Residuals

One problem with the martingale residuals is that they tend to be asymmetric.

A solution is to use **deviance residuals**. For person i, these are defined as a function of the martingale residuals (r_i):

$$\hat{D}_i = \text{sign}(\hat{r}_i)\sqrt{-2[\hat{r}_i + \delta_i log(\delta_i - \hat{r}_i)]}$$

In Stata, the deviance residuals are generated using the same approach as the Cox-Snell residuals.

```
. stcox towdur handling length logcatch, mgale(mg)

. predict devres, deviance
```

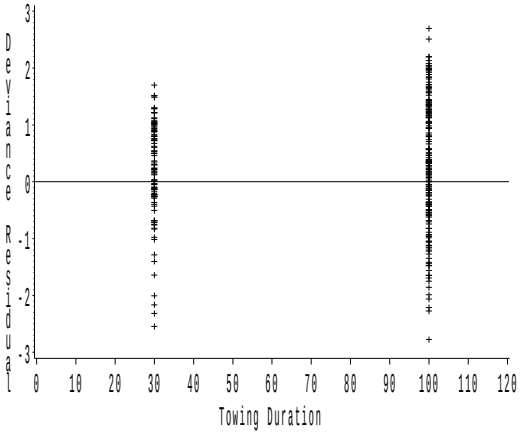
and then they can be plotted versus the predicted log(HR) or the individual covariates, as shown for the Martingale residuals.

In SAS, just use **resdev** option instead of resmart.

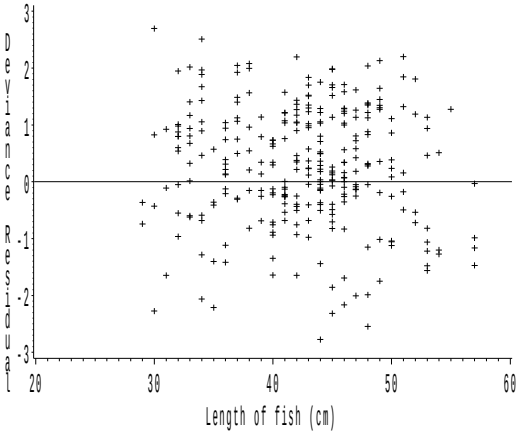
Deviance residuals behave much like residuals from OLS regression (i.e., mean=0, s.d.=1). They are negative for observations with survival times that are smaller than expected.

Deviance Residuals

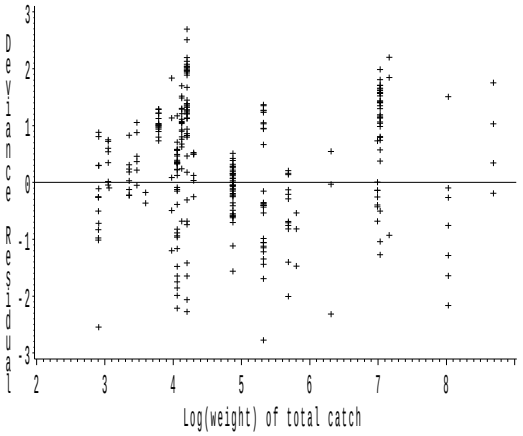
Deviance residuals vs towing duration



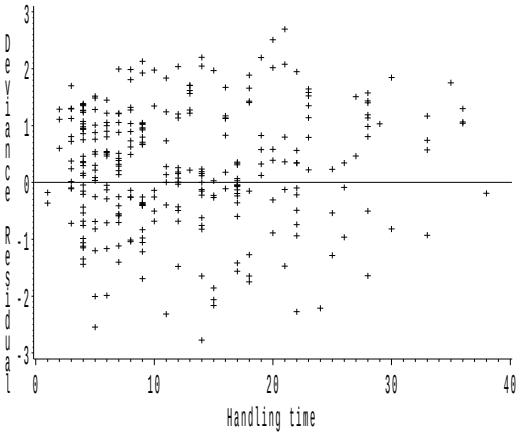
Deviance residuals vs length of fish



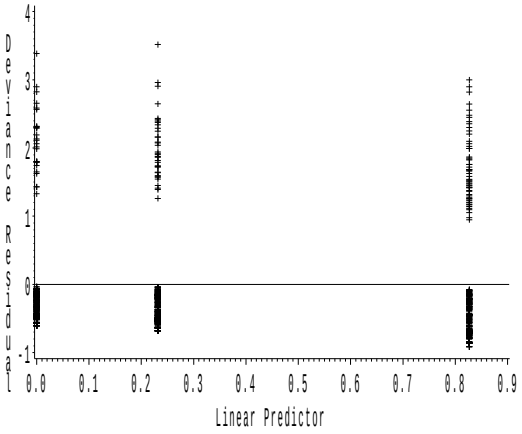
Deviance residuals vs log(catch)



Deviance residuals vs handling



Deviance residuals vs predicted values



(d) Schoenfeld Residuals

These are defined at each observed failure time as:

$$r_{ij}^{s} = Z_{ij}(t_i) - \bar{Z}_j(t_i)$$

Notes:

- represent the difference between the observed covariate and the average over the risk set at that time
- calculated for each covariate
- not defined for censored failure times.
- useful for assessing time trend or lack or proportionality, based on plotting versus event time
- sum to zero, have expected value zero, and are uncorrelated (in large samples)

In Stata, the Schoenfeld residuals are generated in the `stcox` command itself, using the `schoenf(`*`newvar(s)`*`)` option:

```
. stcox towdur handling length logcatch, schoenf(towres handres lenres
logres)

. graph towres survtime
```

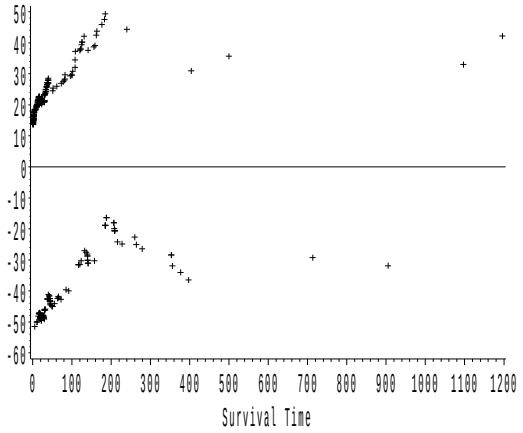
In SAS, add to the output line

```
RESSCH=name1 name2 ... namek
```

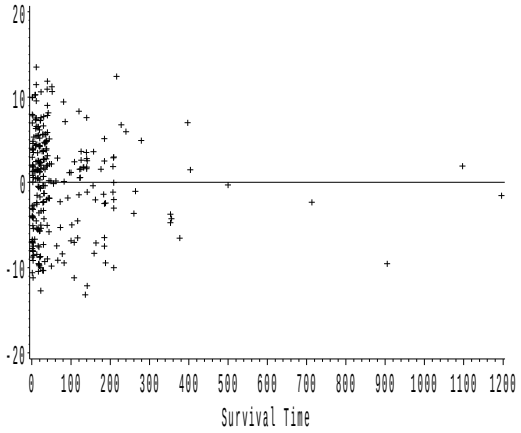
for up to k regressors in the model.

Schoenfeld Residuals

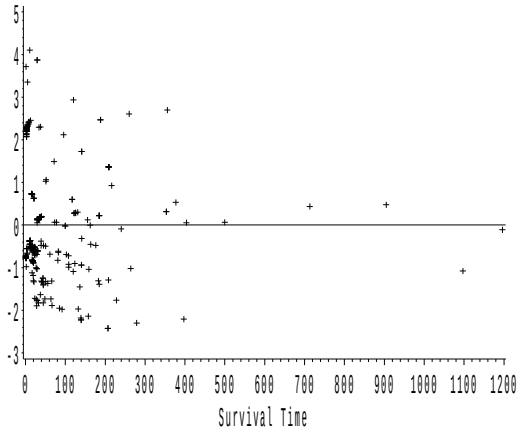
Schoenfeld resids for towing vs survival time



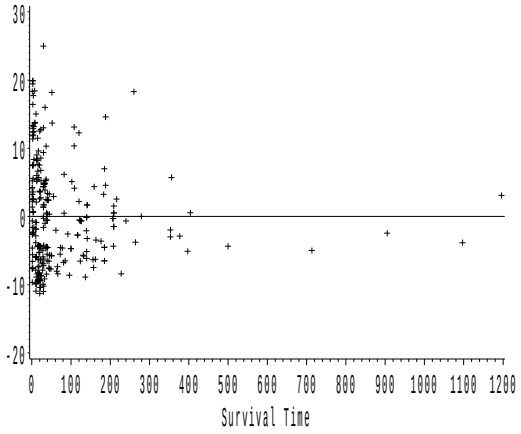
Schoenfeld resids for length vs survival time



Schoenfeld resids for log(catch) vs survival time



Schoenfeld resids for handling vs survival time



(e) Weighted Schoenfeld Residuals

These are actually used more often than the previous unweighted version, because they are more like the typical OLS residuals (i.e., symmetric around 0).

They are defined as:

$$r_{ij}^{w} \;=\; n\widehat{V}\; r_{ij}^{s}$$

where $\widehat{V}$ is the estimated variance of $\hat{\boldsymbol{\beta}}$. The weighted residuals can be used in the same way as the unweighted ones to assess time trends and lack of proportionality.

In Stata, use the command:

```
. stcox towdur length logcatch handling depth, scaledsch(towres2
> lenres2 logres2 handres2 depres2)

. graph logres2 survtime
```

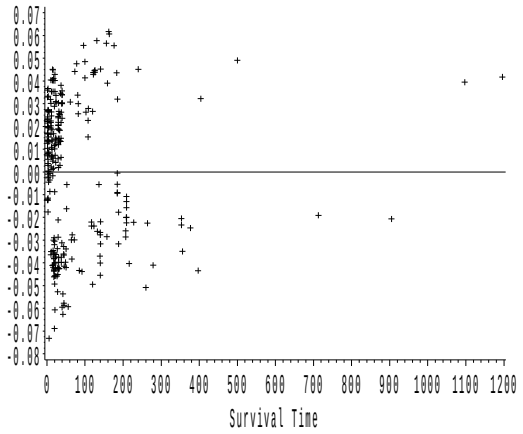
In SAS, add to the output line

```
WTRESSCH=name1 name2 ... namek
```

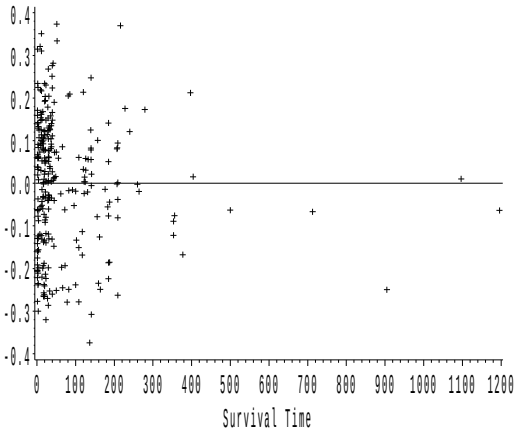
for up to k regressors in the model.

Weighted Schoenfeld Residuals

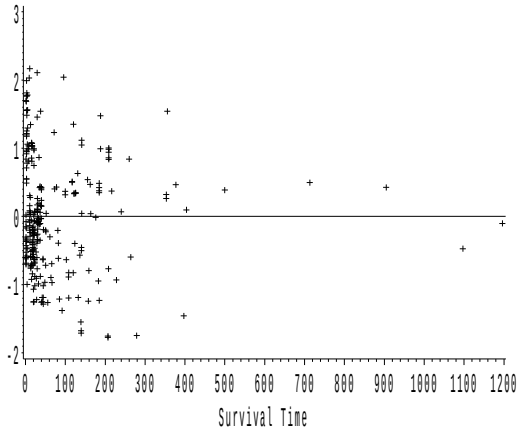
Weighted Schoenfeld resids for towing vs time



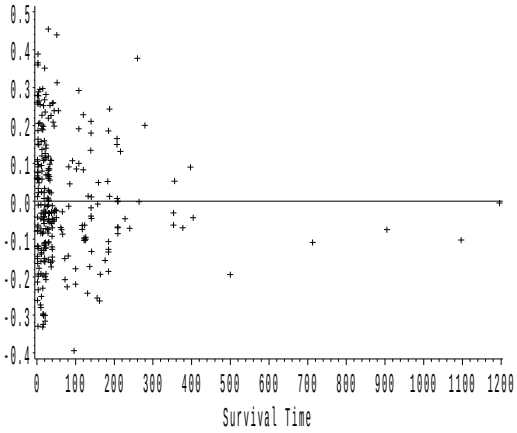
Schoenfeld resids for length vs time



Schoenfeld resids for log(catch) vs time



Schoenfeld resids for handling vs time



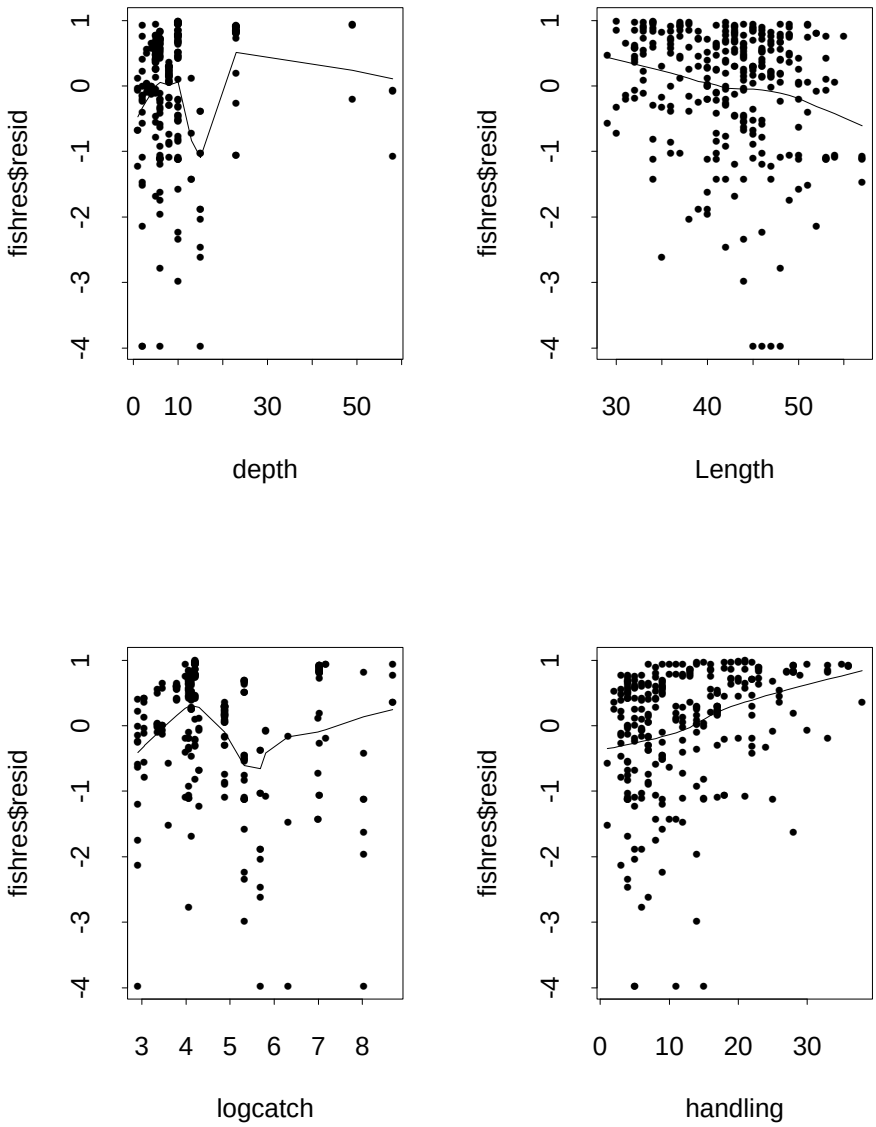
Using Residual plots to explore relationships

If you calculate martingale or deviance residuals without any covariates in the model and then plot against covariates, you obtain a graphical impression of the relationship between the covariate and the hazard.

In Splus, it is easy to do this (also possible in stata using the "estimate" option)

```
** read in the dataset and fit a cox PH model
fish_read.table('fish.data',header=T)
x_fish$towdur
fishres_coxreg(fish$time, fish$censor, x, resid="martingale",iter.max=0)

** the 2 commands below set up the postscript file, with 4 graphs
postscript("fishres.plt",horizontal=F,height=10,width=7)
par(mfrow=c(2,2),oma=c(0,0,2,0))

** plot the martingale residuals vs each of the other covariates
** and add a lowess smoothed fit to the plot
plot(fish$depth, fishres$resid, xlab="depth")
lines(lowess(fish$depth,fishres$resid,iter=0))

plot(fish$length,  fishres$resid, xlab="length")
lines(lowess(fish$length,fishres$resid,iter=0))

plot(fish$handling,  fishres$resid,  xlab="handling")
lines(lowess(fish$handling,fishres$resid,iter=0))

plot(fish$logcatch,  fishres$resid,  xlab="logcatch")
lines(lowess(fish$logcatch,fishres$resid,iter=0))
```

Splus Plots of Martingale Residuals for Cox Model containing only towing duration as a predictor, vs other covariates



(f) Deletion diagnostics

Deletion diagnostics are defined generally as:

$$\delta_i = \hat{\beta} - \hat{\beta}_{(i)}$$

In other words, they are the difference between the estimated regression coefficient using all observations and that without the i-th individual. This can be useful for assessing the **influence** of an individual.

In SAS PROC PHREG, we use the **dfbeta** option:
(Note that there is a separate **dfbeta** calculated for each of the predictors.)

```
proc phreg data=fish;
  model survtime*censor(0)=towdur handling logcatch length;
  id id;
  output out=phinfl dfbeta=dtow dhand dlogc dlength ld=lrchange;

proc univariate data=phinfl;
  var dtow dhand dlogc dlength lrchange;
  id id;
run;
```

The proc univariate procedure will supply the 5 smallest values and the 5 largest values. The "`id`" statement means that these will be labeled with the value of id from the dataset.

(g) Other Influence diagnostics

Other influence diagnostics:

The **LD** option is another method for checking influence. It calculates how much the log-likelihood (x2) would change if the i-th person was removed from the sample.

$$LD_i \;=\; 2\left[logL(\widehat{\boldsymbol{\beta}}) - logL(\widehat{\boldsymbol{\beta}}_{-i})\right]$$

$\widehat{\boldsymbol{\beta}}$ = MLE for all parameters with everyone included
$\widehat{\boldsymbol{\beta}}_{-i}$ = MLE with i-th subject omitted

Again, the proc univariate procedure in SAS will identify the observations with the largest and smallest values of the `lrchange` diagnostic measure.

Can we improve the model?

The plots appear to have some structure, which indicate that we could be leaving something out. It is always a good idea to check for interactions:

In this case, there are several important interactions. I used a backward selection model forcing all main effects to be included, and considering all pairwise interactions. Here are the results:

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
TOWDUR	1	-0.075452	0.01740	18.79679	0.0001	0.927
DEPTH	1	0.123293	0.06400	3.71107	0.0541	1.131
LENGTH	1	-0.077300	0.02551	9.18225	0.0024	0.926
HANDLING	1	0.004798	0.03221	0.02219	0.8816	1.005
LOGCATCH	1	-0.225158	0.07156	9.89924	0.0017	0.798
TOWDEPTH	1	0.002931	0.0004996	34.40781	0.0001	1.003
TOWLNGTH	1	0.001180	0.0003541	11.10036	0.0009	1.001
TOWHAND	1	0.001107	0.0003558	9.67706	0.0019	1.001
DEPLNGTH	1	-0.006034	0.00136	19.77360	0.0001	0.994
DEPHAND	1	-0.004104	0.00118	12.00517	0.0005	0.996

Interpretation:

Handling alone doesn't seem to affect survival, unless it is combined with a longer towing duration or shallower trawling depths.

An alternative modeling strategy when we have fewer covariates

With a dataset with only 5 main effects, it would make sense to consider interactions from the start. How many would there be?

- Fit model with all main effects and pairwise interactions

- Then use backward selection to eliminate non-significant pairwise interactions (remember to force the main effects into the model at this stage)

- Once non-significant pairwise interactions have been eliminated, you could consider backwards selection to eliminate any non-significant main effects that are not involved in remaining interaction terms

- After obtaining final model, use residuals to check fit of model.

Assessing the PH Assumption

So far, we've been considering the following Cox PH model:

$$\begin{aligned} \lambda(t, \mathbf{Z}) &= \lambda_0(t) \ \exp(\boldsymbol{\beta}\mathbf{Z}) \\ &= \lambda_0(t) \ \exp\left(\sum \beta_j Z_j\right) \end{aligned}$$

where β_j is the parameter for the the j-th covariate (Z_j).

Important features of this model:

(1) the baseline hazard depends on t, but not on the covariates $Z_1, ..., Z_p$

(2) the hazard ratio, i.e., $\exp(\boldsymbol{\beta}\mathbf{Z})$, depends on the covariates $\mathbf{Z} = (Z_1, ..., Z_p)$, but not on time t.

Assumption (2) is what led us to call this a proportional hazards model. That's because we could take the ratio of the hazards for two individuals with covariates $\mathbf{Z}_i$ and $\mathbf{Z}_{i'}$, and write it as a constant in terms of the covariates.

Proportional Hazards Assumption

Hazard Ratio:

$$\begin{aligned}
\frac{\lambda(t, \mathbf{Z}_i)}{\lambda(t, \mathbf{Z}_{i'})} &= \frac{\lambda_0(t)\exp(\boldsymbol{\beta}\mathbf{Z}_i)}{\lambda_0(t)\exp(\boldsymbol{\beta}\mathbf{Z}_{i'})} \\
&= \frac{\exp(\boldsymbol{\beta}\mathbf{Z}_i)}{\exp(\boldsymbol{\beta}\mathbf{Z}_{i'})} \\
&= \exp[\boldsymbol{\beta}(\mathbf{Z}_i - \mathbf{Z}_{i'})] \\
&= \exp[\textstyle\sum \beta_j (Z_{ij} - Z_{i'j})] \;=\; \theta
\end{aligned}$$

In the last formula, Z_{ij} is the value of the j-th covariate for the i-th individual. For example, Z_{42} might be the value of GENDER (0 or 1) for the the 4-th person.

We can also write the hazard for the i-th person as a constant times the hazard for the i'-th person:

$$\lambda(t, \mathbf{Z}_i) \;=\; \theta\; \lambda(t, \mathbf{Z}_{i'})$$

Thus, the HR between two types of individuals is constant (i.e., $=\theta$) over time. These are mathematical ways of stating the proportional hazards assumption.

There are several options for checking the assumption of proportional hazards:

I. **Graphical**

   (a) Plots of survival estimates for two subgroups

   (b) Plots of $\log[-\log(\hat{S})]$ vs $\log(t)$ for two subgroups

   (c) Plots of weighted Schoenfeld residuals vs time

   (d) Plots of observed survival probabilities versus expected under PH model (see Kleinbaum, ch.4)

II. **Use of goodness of fit tests** - we can construct a goodness-of-fit test based on comparing the observed survival probability (from `sts list`) with the expected (from `stcox`) under the assumption of proportional hazards - see Kleinbaum ch.4

III. **Including interaction terms between a covariate and** t (time-dependent covariates)

How do we interpret the above?

Kleinbaum (and other texts) suggest a strategy of assuming that PH holds unless there is very strong evidence to counter this assumption:

- estimated survival curves are fairly separated, then cross
- estimated log cumulative hazard curves cross, or look very unparallel over time
- weighted Schoenfeld residuals clearly increase or decrease over time (you could fit a OLS regression line and see if the slope is significant)
- test for time $\times$ covariate interaction term is significant (this relates to time-dependent covariates)

If PH doesn't exactly hold for a particular covariate but we fit the PH model anyway, then what we are getting is sort of an average HR, averaged over the event times.

In most cases, this is not such a bad estimate. Allison claims that too much emphasis is put on testing the PH assumption, and not enough to other important aspects of the model.

Implications of proportional hazards

Consider a PH model with a single covariate, Z:

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

What does this imply for the relation between the survivorship functions at various values of Z?

Under PH,

$$\log[-\log[S(t; Z)]] = \log[-\log[S_0(t)]] + \beta Z$$

In general, we have the following relationship:

$$\begin{aligned}
\Lambda_i(t) &= \int_0^t \lambda_i(u)du \\
&= \int_0^t \lambda_0(u)\exp(\boldsymbol{\beta}\mathbf{Z}_i)du \\
&= \exp(\boldsymbol{\beta}\mathbf{Z}_i)\int_0^t \lambda_0(u)du \\
&= \exp(\boldsymbol{\beta}\mathbf{Z}_i)\,\Lambda_0(t)
\end{aligned}$$

This means that the ratio of the cumulative hazards is the same as the ratio of hazard rates:

$$\frac{\Lambda_i(t)}{\Lambda_0(t)} = \exp(\boldsymbol{\beta}\mathbf{Z}_i) = \exp(\beta_1 Z_{1i} + \cdots + \beta_p Z_{pi})$$

Using the above relationship, we can show that:

$$\begin{aligned} \boldsymbol{\beta}\mathbf{Z}_i &= \log\left(\frac{\Lambda_i(t)}{\Lambda_0(t)}\right) \\ &= \log\Lambda_i(t) - \log\Lambda_0(t) \\ &= \log[-\log S_i(t)] - \log[-\log S_0(t)] \\ \text{so} \quad \log[-\log S_i(t)] &= \log[-\log S_0(t)] + \boldsymbol{\beta}\mathbf{Z}_i \end{aligned}$$

Thus, to assess if the hazards are actually proportional to each other over time (using graphical option I(b))

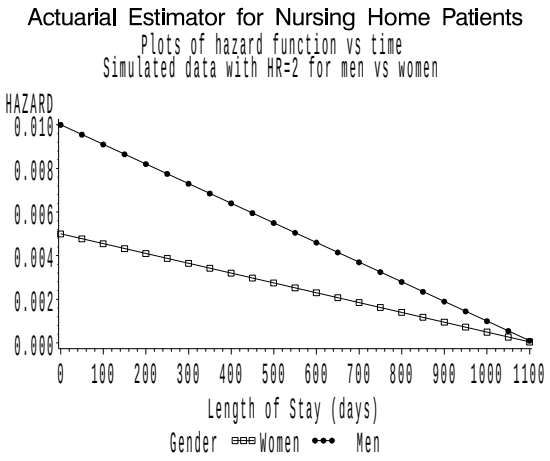
- calculate Kaplan Meier Curves for various levels of Z
- compute $\log[-\log(\hat{S}(t;Z))]$ (i.e., log cumulative hazard)
- plot vs log-time to see if they are parallel (lines or curves)

Note: If Z is continuous, break into categories.

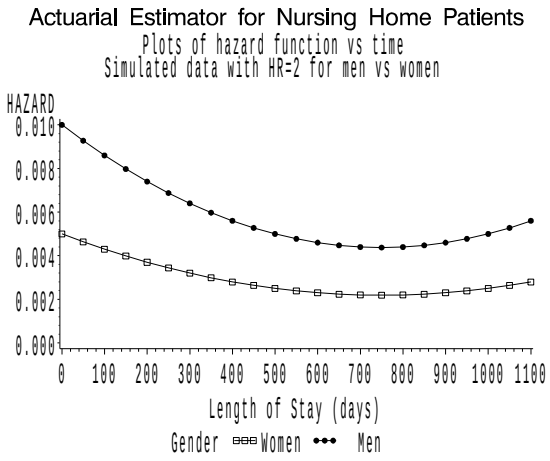
Question: Why not just compare the underlying hazard rates to see if they are proportional?

Here's two simulated examples with hazards which are truly proportional between the two groups:

Weibull-type hazard:



U-shaped hazard:

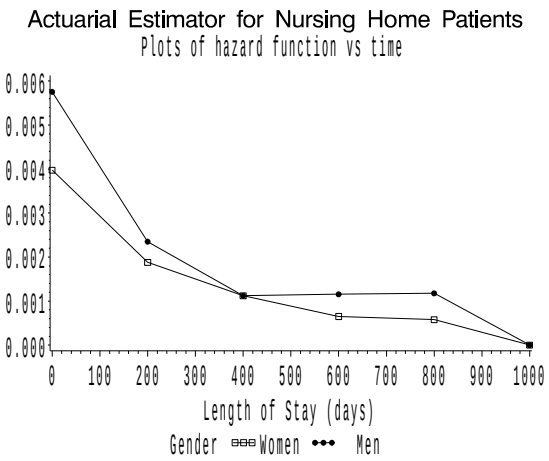


Reason 1: It's hard to eyeball these figures and see that the hazard rates are proportional - it would be easier to look for a constant shift between lines.

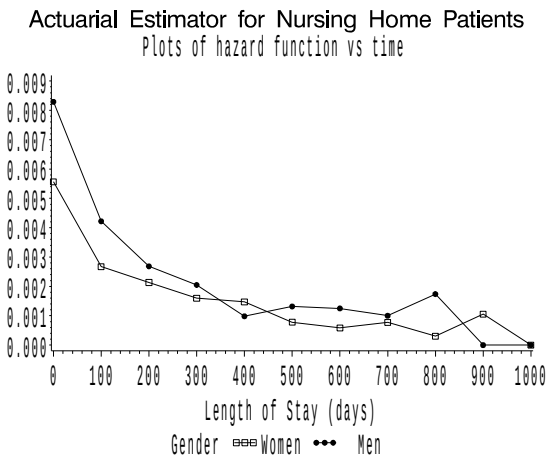
Reason 2: Estimated hazard rates tend to be more unstable than the cumulative hazard rate

Consider the nursing home example (where we think PH is reasonable). If we group the data into intervals and calculate the hazard rate using actuarial method, we get these plots:

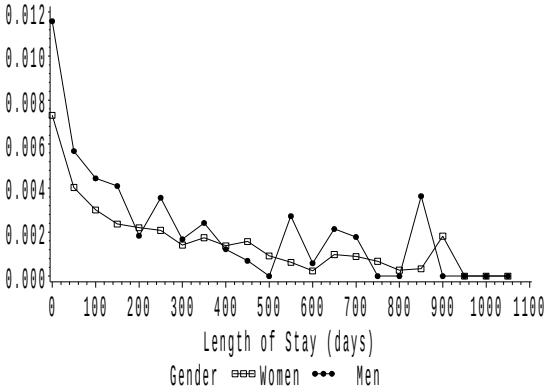
200 day intervals:



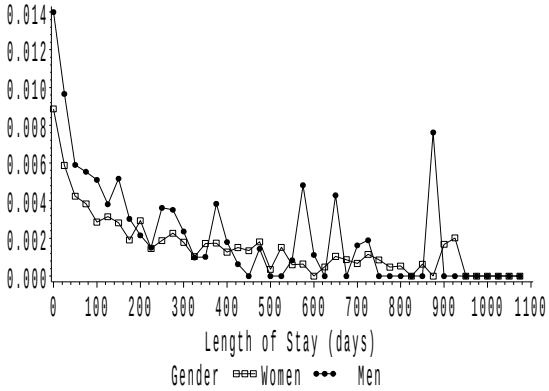
100 day intervals:



50 day intervals:



25 day intervals:



In contrast, the log cumulative hazard plots are easier to interpret and tend to give more stable estimates

Ex: Nursing Home - gender and marital status

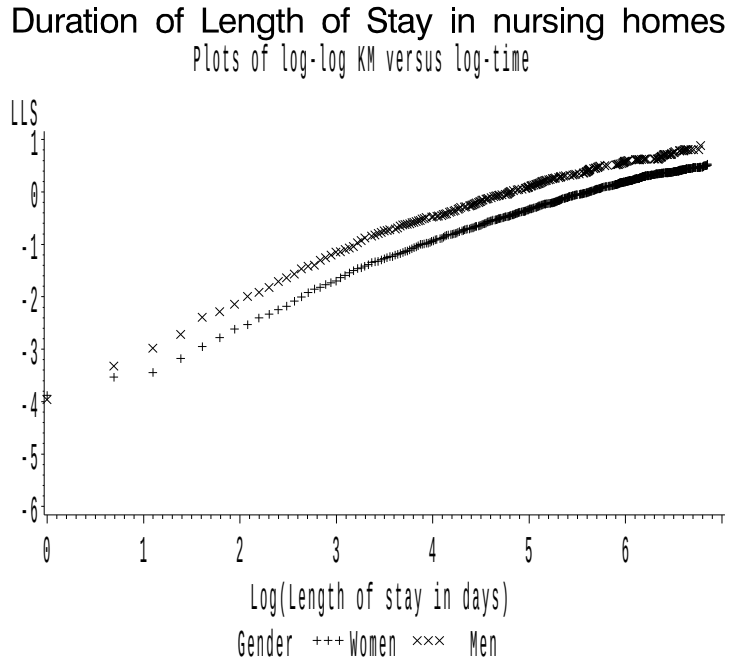
```
proc lifetest data=pop outsurv=survres;
  time los*fail(0);
  strata gender;
  format gender sexfmt.;
  title 'Duration of Length of Stay in nursing homes';

data survres;
  set survres;
  label log_los='Log(Length of stay in days)';
  if los > 0 then log_los=log(los);
  if survival<1 then lls=log(-log(survival));

proc gplot data=survres;
  plot lls*log_los=gender;
  format gender sexfmt.;
  title2 'Plots of log-log KM versus log-time';
run;
```

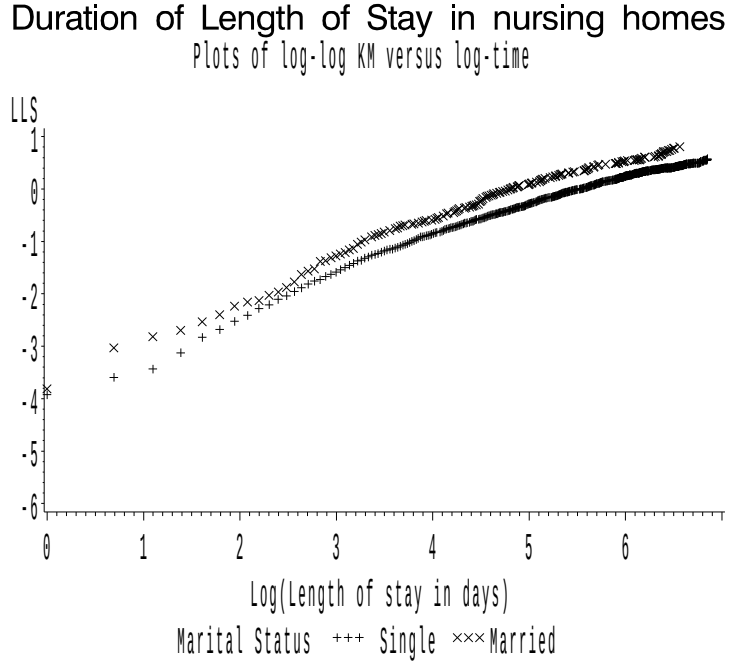
The statements for marital status are similar, substituting MARRIED for GENDER.

Note: This is equivalent to comparing plots of the log cumulative hazard, $\log(\hat{\Lambda}(t))$, between the covariate levels, since

$$\Lambda(t) = \int_0^t \lambda(u; Z)du = -\log[S(t)]$$

Assessment of proportional hazards for gender and marital status in nursing home data (Morris)





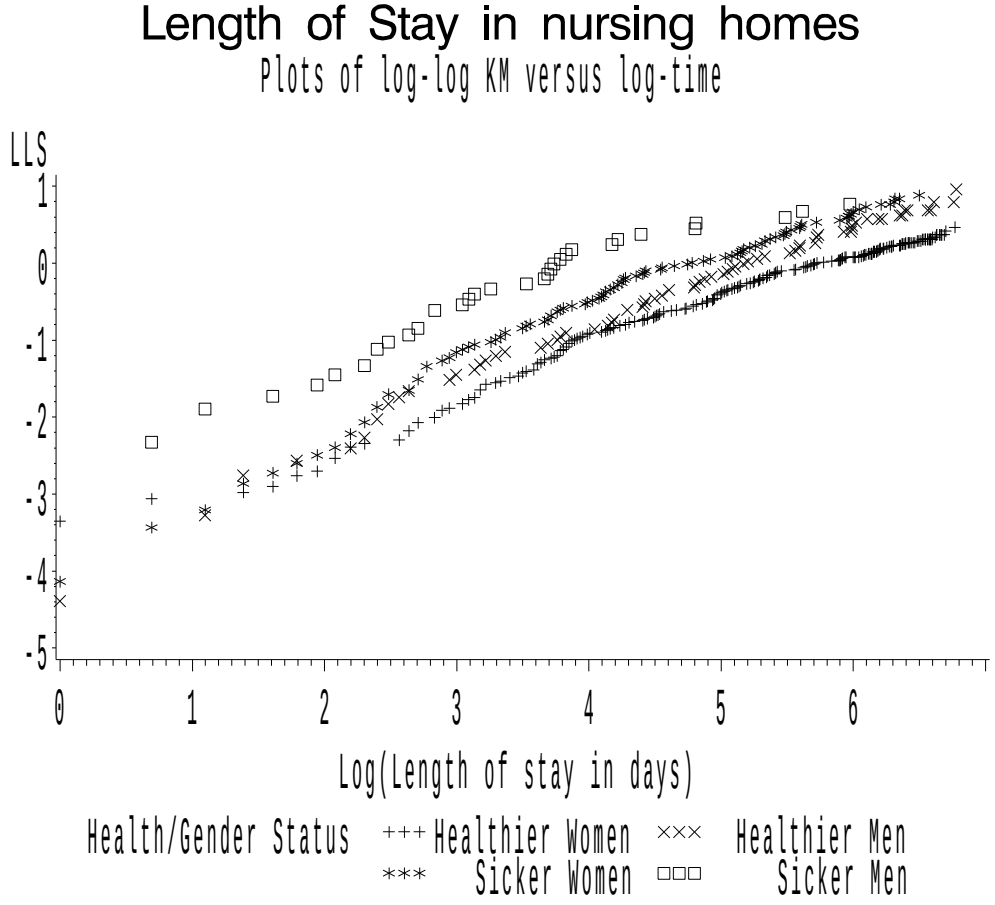
Assessing proportionality with several covariates

If there is enough data and you only have a couple of covariates, create a new covariate that takes a different value for every combination of covariate values.

Example: Health status and gender for nursing home

```
data pop;
  infile 'ch12.dat';
  input los age rx gender married health fail;
  if gender=0 and health=2 then hlthsex=1;
  if gender=1 and health=2 then hlthsex=2;
  if gender=0 and health=5 then hlthsex=3;
  if gender=1 and health=5 then hlthsex=4;

proc format;
  value hsfmt
    1='Healthier Women'
    2='Healthier Men'
    3='Sicker Women'
    4='Sicker Men';

proc lifetest data=pop outsurv=survres;
  time los*fail(0);
  strata hlthsex;
  format hlthsex hsfmt.;
  title 'Length of Stay in nursing homes';

data survres;
  set survres;
  label log_los='Log(Length of stay in days)';
  label hlthsex='Health/Gender Status';
  if los > 0 then log_los=log(los);
  if survival<1 lls=log(-log(survival));

proc gplot data=survres;
  plot lls*log_los=hlthsex;
  format hlthsex hsfmt.;
  title2 'Plots of log-log KM versus log-time';
run;
```

Log[-log(survival)] Plots for Health status*gender



If there are too many covariates (or not enough data) for this, then there is a way to test proportionality for each variable, one at a time, using the stratification option.

What if proportional hazards fails?

- do a stratified analysis
- include a time-varying covariate to allow changing hazard ratios over time
- include interactions with time

The second two options relate to time-dependent covariates, which will be covered in future lectures.

We will focus on the first alternative, and then the second two options will be briefly described.

Stratified Analyses

Suppose:

- we are happy with the proportionality assumption on Z_1
- proportionality simply does not hold between various levels of a second variable Z_2.

If Z_2 is discrete (with a levels) and there is enough data, fit the following **stratified model**:

$$\lambda(t; Z_1, Z_2) = \lambda_{Z_2}(t)e^{\beta Z_1}$$

For example, a new treatment might lead to a 50% decrease in hazard of death versus the standard treatment, but the hazard for standard treatment might be different for each hospital.

A stratified model can be useful both for primary analysis and for checking the PH assumption.

Assessing PH Assumption for Several Covariates

Suppose we have several covariates ($\mathbf{Z} = Z_1, Z_2, \ldots\ Z_p$), and we want to know if the following PH model holds:

$$\lambda(t; \mathbf{Z}) = \lambda_0(t)\ e^{\beta_1 Z_1 + \ldots + \beta_p Z_p}$$

To start, we fit a model which stratifies by Z_k:

$$\lambda(t; \mathbf{Z}) = \lambda_{0Z_k}(t)\ e^{\beta_1 Z_1 + \ldots + \beta_{k-1} Z_{k-1} + \beta_{k+1} Z_{k+1} + \ldots + \beta_p Z_p}$$

Since we can estimate the survival function for any subgroup, we can use this to estimate the baseline survival function, $S_{0Z_k}(t)$, for each level of Z_k.

Then we compute $-\log S(t)$ for each level of Z_k, controlling for the other covariates in the model, and graphically check whether the log cumulative hazards are parallel across strata levels.

Ex: PH assumption for `gender` (nursing home data):

- include `married` and `health` as covariates in a Cox PH model, but *stratify* by `gender`.
- calculate the baseline survival function for each level of the variable `gender` (i.e., males and females)
- plot the log-cumulative hazards for males and females and evaluate whether the lines (curves) are parallel

In the above example, we make the PH assumption for `married` and `health`, but not for `gender`.

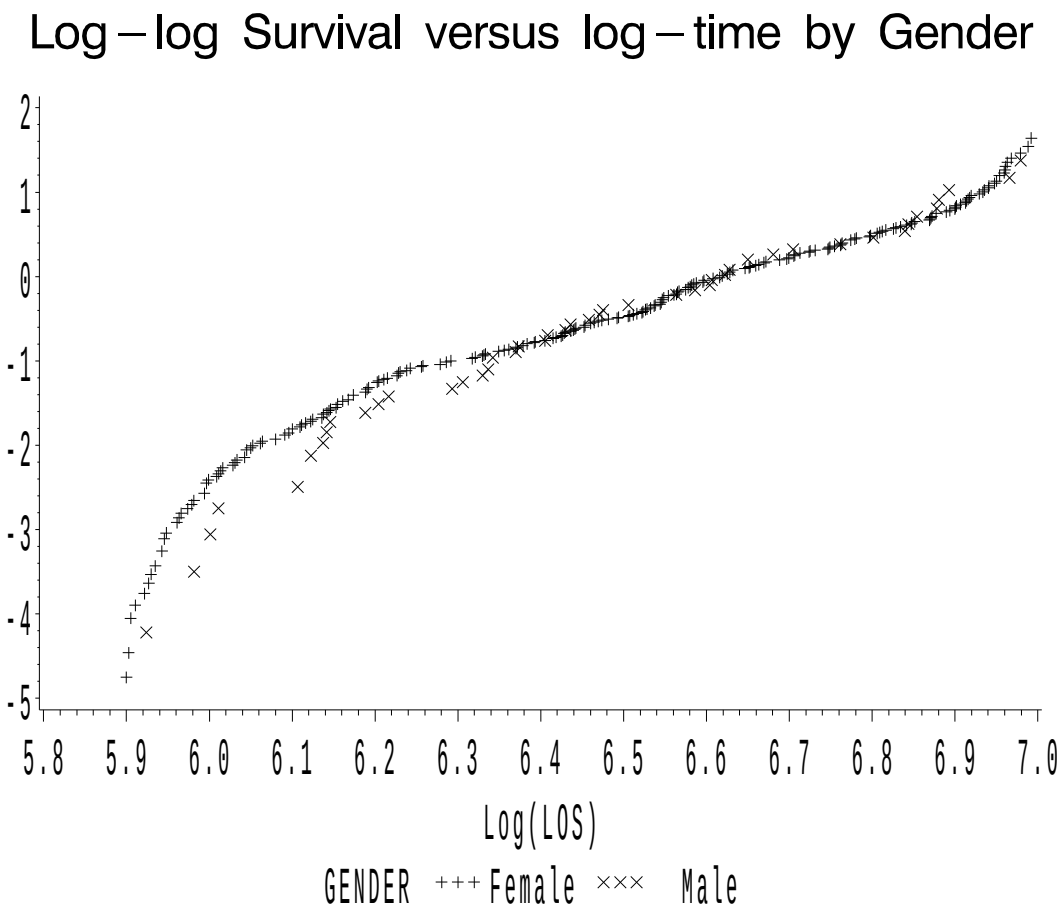
This is like getting a KM survival estimate for each gender without assuming PH, but is more flexible since we can control for other covariates.

We would repeat the stratification for each variable for which we wanted to check the PH assumption.

SAS Code for Assessing PH within Stratified Model:

```
data pop;
  infile 'ch12.dat';
  input los age rx gender married health fail;
  if los<=0 then delete;

data inrisks;
  input married health;
  cards;
0 2
;

proc format;
  value sexfmt
    1='Male'
    0='Female';

proc phreg data=pop;
  model los*fail(0)=married health;
  strata gender;
  baseline covariates=inrisks out=outph
      loglogs=lls / nomean;

proc print data=outph;
title 'Log Cumulative Hazard Estimates by Gender';
title2 'Controlling for Marital and Health Status';

data outph;
  set outph;
  if los>0 then log_los=log(los);
  label log_los='Log(LOS)'
        lls='Log Cumulative Hazard';

proc gplot data=outph;
  plot lls*log_los=gender;
  format gender sexfmt.;
  title1 'Log-log Survival versus log-time by Gender';
run;
```

Log[-log(survival)] Plots for Gender Controlling for Marital and Health Status



Models with Time-dependent Interactions

Consider a PH model with two covariates Z_1 and Z_2. The standard PH model assumes

$$\lambda(t; Z) = \lambda_0(t)\ e^{\beta_1 Z_1 + \beta_2 Z_2}$$

However, if the log-hazards are not really parallel between the groups defined by Z_2, then you can try adding an interaction with time:

$$\lambda(t; Z) = \lambda_0(t)\ e^{\beta_1 Z_1 + \beta_2 Z_2 + \beta_3 Z_2 * t}$$

A test of the coefficient β_3 would be a test of the proportional hazards assumption for Z_2.

If β_3 is positive, then the hazard ratio would be increasing over time; if negative, then decreasing over time.

Changes in covariate status sometimes occur naturally during a study (ex. patient gets a kidney transplant), and are handled by introducing *time-dependent covariates*.

Using STATA to Assess Proportional Hazards

Stata has two commands which can be used to graphically assess the proportional hazards assumption, using graphical options (b) and (d) described previously:

- `stphplot:` plots $-\log[-\log(-(S(t))]$ curves for each category of a nominal or ordinal independent variable versus log(time). Optionally, these estimates can be adjusted for other covariates.

- `stcoxkm:` plots Kaplan-Meier observed survival curves and compares them to the Cox predicted curves for the same variable. (No need to run stcox prior to this command, it will be done automatically)

For either command, you must have `stset` your data first.

You must specify `by()` with `stcoxkm` and you must specify either `by()` or `strata()` with `stphplot`.

Assessing PH Assumption for a Single Covariate by Comparing $-\log[-\log(S(t))]$ Curves

```
. use nurshome

. stset los fail

. stphplot, by(gender)
```

Note that the lines will be going from top left to bottom right, rather than bottom left to top right, since we are plotting $-\log[-\log(S(t))]$ rather than $\log[-\log(S(t))]$.

This will give a plot similar to that on p.10 (top).

Of course, you'll want to make your plot prettier by adding titles and labels, as follows:

```
. stphplot, by(gender)  xlab ylab b2(log(Length of Stay))
>title(Evaluation of PH Assumption) saving(phplot)
```

Assessing PH Assumption for Several Covariates by Comparing $-\log[-\log(S(t))]$ Curves

```
. use nurshome

. stset los fail

. gen hlthsex=1

. replace hlthsex=2 if health==2 & gender==1

. replace hlthsex=3 if health==5 & gender==0

. replace hlthsex=4 if health==5 & gender==1

. tab hlthsex

. stphplot, by(hlthsex)
```

This will give a plot similar to that on p.12.

Assessing PH Assumption for a Single Covariate Controlling for the Levels of Other Covariates

```
. use nurshome

. stset los fail

. stphplot, strata(gender) adjust(married health)
```

This will produce a plot similar to that on p.18.

Assessing PH Assumption for a Covariate By Comparing Cox PH Survival to KM Survival

To construct plots based on option I(d), use the `stcoxkm` command, either for a single covariate or for a newly generated covariate (like `hlthsex`) which represents combined levels of more than one covariate.

```
. use nurshome

. stset los fail

. stcoxkm, by(gender)

. stcoxkm, by(hlthsex)
```

As usual, you'll want to add titles, labels, and save your graph for later use.

Time varying (or time-dependent) covariates

References:

Allison (*)	p.138-153
Hosmer & Lemeshow	Chapter 7, Section 3
Kalbfleisch & Prentice	Section 5.3
Collett	Chapter 7
Kleinbaum	Chapter 6
Cox & Oakes	Chapter 8
Andersen & Gill	Page 168 (Advanced!)

So far, we've been considering the following Cox PH model:

$$\begin{aligned} \lambda(t, \mathbf{Z}) &= \lambda_0(t)\ \exp(\boldsymbol{\beta}\mathbf{Z}) \\ &= \lambda_0(t) \exp(\textstyle\sum \beta_j Z_j) \end{aligned}$$

where β_j is the parameter for the the j-th covariate (Z_j).

Important features of this model:

(1) the baseline hazard depends on t, but not on the covariates $Z_1, ..., Z_p$

(2) the hazard ratio $\exp(\boldsymbol{\beta}\mathbf{Z})$ depends on the covariates $Z_1, ..., Z_p$, but not on time t.

Now we want to relax the second assumption, and allow the hazard ratio to depend on time t.

Example to motivate time-dependent covariates

Stanford Heart transplant example:

Variables:

- SURVIVAL - time from program enrollment until death or censoring
- DEAD - indicator of death (1) or censoring (0)
- TRANSPL - whether patient ever had transplant
  (1 if yes, 2 if no)
- SURGERY - previous heart surgery prior to program
- AGE - age at time of acceptance into program
- WAIT - time from acceptance into program until transplant surgery (=. for those without transplant)

Initially, a Cox PH model was fit for predicting survival time:

$$\lambda(t, \mathbf{Z}) \;=\; \lambda_0(t)\ \exp(\beta_1 * transpl + \beta_2 * surgery + \beta_3 * age)$$

However, this model could give misleading results, since patients who died more quickly had less time available to get transplants. A model with a time dependent indicator of whether a patient had a transplant at each point in time might be more appropriate:

$$\lambda(t, \mathbf{Z}) \;=\; \lambda_0(t)\ \exp(\beta_1 * trnstime + \beta_2 * surgery + \beta_3 * age)$$

where TRNSTIME = 1 if TRANSPL=1 and WAIT$> t$

SAS code for these two models

Time-independent covariate for TRANSPL:

```
proc phreg data=stanford;
  model survival*dead(0)=transpl surgery age;
run;
```

Time-dependent covariate for TRANSPL:

```
proc phreg data=stanford;
  model survival*dead(0)=trnstime surgery age;
  if wait>survival or wait=. then trnstime=0;
    else trnstime=1;
run;
```

If we add time-dependent covariates or interactions with time to the Cox proportional hazards model, then it is not "proportional hazards" model any longer.

We refer to it as an "extended Cox model".

Comparison with a single binary predictor (like heart transplant):

- A standard Cox PH model would compare the survival distributions between those without a transplant (ever) to those with a transplant. A subject's transplant status at the end of the study would determine which category they were put into for the entire study follow-up.
- An extended Cox model would compare the risk of an event between transplant and non-transplant at each event time, but would re-evaluate which risk group each person belonged in based on whether they'd had a transplant by that time.

Recidivism Example: (see Allison, p.42)

Recidivism study:

432 male inmates were followed for one year after release from prison, to evaluate risk of re-arrest as function of financial aid (FIN), age at release (AGE), race (RACE), full-time work experience prior to first arrest (WEXP), marital status (MAR), parole status (PARO=1 if released with parole, 0 otherwise), and number of prior convictions (PRIO). Data were also collected on employment status over time during the year.

Time-independent model:

A time independent model might include the employment status of the individual at the beginning of the study (1 if employed, 0 if unemployed), or perhaps at any point during the year.

Time-dependent model:

However, employment status changes over time, and it may be the more recent employment status that would affect the hazard for re-arrest. For example, we might want to define a time-dependent covariate for each month of the study that indicates whether the individual was employed during the past month.

Extended Cox Model

Framework:

For individual i, suppose we have their failure time, failure indicator, and a summary of their covariate values over time:

$$(X_i, \delta_i, \{Z_i(t), t \in [0, X_i]\}),$$

$\{Z_i(t), t \in [0, X_i]\}$ represents the **covariate path** for the i-th individual while they are in the study, and the covariates can take different values at different times.

Assumptions:

- conditional on an individual's covariate history, the hazard for failure at time t depends only on the value of the covariates **at that time**:

$$\lambda(t; \{Z_i(u), u \in [0, t]\}) = \lambda(t; Z_i(t))$$

- the Cox model for the hazard holds:

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

Survivor function:

$$S(t; Z) = \exp\{-\int_0^t \exp(\beta Z(u))\ \lambda_0(u) du\}$$

and depends on the values of the time dependent variables over the interval from 0 to t.

This is the classic formulation of the time varying Cox regression survival model.

Kinds of time-varying covariates:

- **internal covariates:**

  variables that relate to the individuals, and can only be measured when an individual is alive, e.g. white blood cell count, CD4 count

- **external covariates:**
  - variable which changes in a known way, e.g. age, dose of drug
  - variable that exists totally independently of all individuals, e.g. air temperature

Applications and Examples

The extended Cox model is used:

I. When **important covariates change** during a study

- **Framingham Heart study**

  5209 subjects followed since 1948 to examine relationship between risk factors and cardiovascular disease. A particular example:

  Outcome: time to congestive heart failure
  Predictors: age, systolic blood pressure, # cigarettes per day

- **Liver Cirrhosis** (Andersen and Gill, p.528)

  Clinical trial comparing treatment to placebo for cirrhosis. The outcome of interest is time to death. Patients were seen at the clinic after 3, 6 and 12 months, then yearly.

  Fixed covariates: treatment, gender, age (at diagnosis)
  Time-varying covariates: alcohol consumption, nutritional status, bleeding, albumin, bilirubin, alkaline phosphatase and prothrombin.

- **Recidivism Study** (Allison, p.42)

II. For **cross-over studies**, to indicate change in treatment

- **Stanford heart study** (Cox and Oakes p.129)
  Between 1967 and 1980, 249 patients entered a program at Stanford University where they were registered to receive a heart transplant. Of these, 184 received transplants, 57 died while waiting, and 8 dropped out of the program for other reasons. Does getting a heart transplant improve survival? Here is a sample of the data:

Waiting time	transplant?	survival post transplant	total survival	final status
49	2	.	.	1
5	2	.	.	1
0	1	15	15	1
35	1	3	38	1
17	2	.	.	1
11	1	46	57	1

etc

(survival is not indicated above for those without transplants, but was available in the dataset)

Naive approach: Compare the total survival of transplanted and non-transplanted.

Problem: Length Bias!

III. For **Competing Risks** Analysis

For example, in cancer clinical trials, "tumor response" (or shrinking of the tumor) is used as an outcome. However, clinicians want to know whether tumor response correlates with survival.

For this purpose, we can fit an extended Cox model for time to death, with tumor response as a time dependent covariate.

IV. For **testing the PH assumption**

For example, we can fit these two models:

(1) **Time independent covariate** Z_1

$$\lambda(t, \mathbf{Z}) \;=\; \lambda_0(t)\;\exp(\beta_1 * Z_1)$$

The hazard ratio for Z_1 is $\exp(\beta_1)$.

(2) **Time dependent covariate** Z_1

$$\lambda(t, \mathbf{Z}) \;=\; \lambda_0(t)\;\exp(\beta_1 * Z_1 + \beta_2 * Z_1 * t)$$

The hazard ratio for Z_1 is $\exp(\beta_1 + \beta_2 t)$.

(note: we may want to replace t by $(t - t_0)$, so that $\exp(\beta_1)$ represents HR at some convenient time, like the median survival time.)

A test of the parameter β_2 is a test of the PH assumption.

(how do we get the test? ...using the Wald test from the output of second model, or LR test formed by comparing the log-likelihoods of the two models)

Partial likelihood with time-varying covariates

Starting out just as before...

Suppose there are K distinct failure (or death) times, and let $(\tau_1, ...\tau_K)$ represent the K ordered, distinct death times. For now, assume there are no tied death times.

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

Failure: Let i_j denote the label or identity of the individual who fails at time τ_j, including the value of their time-varying covariate during their time in the study

$$\{Z_{i_j}(t), t \in [0, \tau_j]\}$$

History: Let H_j denote the "history" of the entire data set, up to the j-th death or failure time, including the time of the failure, but not the identity of the one who fails, also including the values of all covariates for everyone up to and including time τ_j.

Partial Likelihood: We have seen previously that the partial likelihood can be written as

$$\begin{aligned} L(\beta) &= \prod_{j=1}^{d} P(i_j | H_j) \\ &= \prod_{j=1}^{d} \frac{\lambda(\tau_j; \mathbf{Z}_j(\tau_j))}{\Sigma_{\ell \in \mathcal{R}(\tau_j)} \lambda(\tau_j; \mathbf{Z}_\ell(\tau_j))} \end{aligned}$$

Under the PH assumption, this is:

$$L(\boldsymbol{\beta}) \;=\; \prod_{j=1}^{d} \frac{\exp(\boldsymbol{\beta}\mathbf{Z}_{jj})}{\Sigma_{\ell \in \mathcal{R}(\tau_j)} \exp(\boldsymbol{\beta}\mathbf{Z}_{\ell j})}$$

where $\mathbf{Z}_{\ell j}$ is a short-cut way to denote the value of the covariate vector for the ℓ-th person at the j-th death time, ie:

$$\mathbf{Z}_{\ell j} = \mathbf{Z}_{\ell}(\tau_j)$$

What if Z is not measured for person ℓ at time τ_j?

- use the most recent value (assumes step function)
- interpolate
- impute based on some model

Inference (i.e. estimating the regression coefficients, constructing score tests, etc.) proceeds similarly to standard case. The main difference is that the values of Z will change at each risk set.

Allison notes that it is very easy to write down a Cox model with time-dependent covariates, but much harder to fit (computationally) and interpret.

Old Example revisited:

Group 0: $4^+, 7, 8^+, 9, 10^+$

Group 1: $3, 5, 5^+, 6, 8^+$

Let Z_1 be group, and add another fixed covariate Z_2

ID	fail	censor	Z_1	Z_2	$e^{(\beta_1 Z_1 + \beta_2 Z_2)}$
1	3	1	1	1	$e^{\beta_1+\beta_2}$
2	4	0	0	1	e^{β_2}
3	5	1	1	1	$e^{\beta_1+\beta_2}$
4	5	0	1	0	e^{β_1}
5	6	1	1	1	$e^{\beta_1+\beta_2}$
6	7	1	0	0	1
7	8	0	0	1	e^{β_2}
8	8	0	1	0	e^{β_1}
9	9	1	0	1	e^{β_2}
10	10	0	0	0	1

ordered failure time (τ_j)	Individuals at risk	failure ID	Partial Likelihood contribution
3			
5			
6			
7			
9			

Example continued:

Now suppose Z_2 (a completely different covariate) is a time varying covariate:

				$Z_2(t)$						
ID	fail	censor	Z_1	3	4	5	6	7	8	9
1	3	1	1	0						
2	4	0	0	1	1					
3	5	1	1	1	1	1				
4	5	0	1	0	0	0				
5	6	1	1	0	0	0	0			
6	7	1	0	0	0	0	1	1		
7	8	0	0	0	0	0	0	0	0	
8	8	0	1	0	0	0	0	1	1	
9	9	1	0	0	0	0	1	1	1	1
10	10	0	0	0	1	1	1	1	1	1

ordered failure time (τ_j)	Individuals at risk	failure ID	Partial Likelihood contribution
3			
5			
6			
7			
9			

SAS solution to previous examples

```
Title 'Ph regression:  small class example';
data ph;
    input time status group z3 z4 z5 z6 z7 z8 z9;
    cards;
3   1   1  0  .  .  .  .  .  .
4   0   0  1  1  .  .  .  .  .
5   1   1  1  1  1  .  .  .  .
5   0   1  0  0  0  .  .  .  .
6   1   1  0  0  0  0  .  .  .
7   1   0  0  0  0  1  1  .  .
8   0   0  0  0  0  0  0  0  .
8   0   1  0  0  0  0  1  1  .
9   1   0  0  0  0  1  1  1  1
10  0   0  0  1  1  1  1  1  1
run;


proc phreg ;
  model time*status(0)=group z3 ;
run;

proc phreg ;
  model time*status(0)=group z ;
  z=z3;
  if (time >= 4) then z=z4;
  if (time >= 5) then z=z5;
  if (time >= 6) then z=z6;
  if (time >= 7) then z=z7;
  if (time >= 8) then z=z8;
  if (time >= 9) then z=z9;
run;
```

SAS output from fitting both models

Model with z3:

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	16.953	13.699	3.254 with 2 DF (p=0.1965)
Score	.	.	3.669 with 2 DF (p=0.1597)
Wald	.	.	2.927 with 2 DF (p=0.2315)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
GROUP	1	1.610529	1.21521	1.75644	0.1851	5.005
Z3	1	1.360533	1.42009	0.91788	0.3380	3.898

Model with time-dependent Z:

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	16.953	14.226	2.727 with 2 DF (p=0.2558)
Score	.	.	2.725 with 2 DF (p=0.2560)
Wald	.	.	2.271 with 2 DF (p=0.3212)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
GROUP	1	1.826757	1.22863	2.21066	0.1371	6.214
Z	1	0.705963	1.20630	0.34249	0.5584	2.026

The Stanford Heart Transplant data

```
Title 'Stanford heart transplant data: C & O Table 8.1';

data heart;
  infile 'heart.dat';
  input wait trans post surv status ;
run;

data heart;
  set heart;
  if trans=2 then surv=wait;
run;

*** naive analysis;
proc phreg;
  model surv*status(2)=tstat;
  tstat=2-trans;

*** analysis with time-dependent covariate;
proc phreg;
  model surv*status(2)=tstat;
  tstat = 0;
  if (trans=1 and surv >= wait) then tstat = 1;
run;
```

The second model took about twice as long to run as the first model, which is usually the case for models with time-dependent covariates.

RESULTS for Stanford Heart Transplant data:

Naive model with fixed transplant indicator:

Criterion	Covariates	Covariates	Model Chi-Square
-2 LOG L	718.896	674.699	44.198 with 1 DF (p=0.0001)
Score	.	.	68.194 with 1 DF (p=0.0001)
Wald	.	.	51.720 with 1 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
TSTAT	1	-1.999356	0.27801	51.72039	0.0001	0.135

Model with time-dependent transplant indicator:

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	1330.220	1312.710	17.510 with 1 DF (p=0.0001)
Score	.	.	17.740 with 1 DF (p=0.0001)
Wald	.	.	17.151 with 1 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
TSTAT	1	-0.965605	0.23316	17.15084	0.0001	0.381

Recidivism Example:

Hazard for arrest within one year of release from prison:

Model without employment status

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	1350.751	1317.496	33.266 with 7 DF (p=0.0001)
Score	.	.	33.529 with 7 DF (p=0.0001)
Wald	.	.	32.113 with 7 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
FIN	1	-0.379422	0.1914	3.931	0.0474	0.684
AGE	1	-0.057438	0.0220	6.817	0.0090	0.944
RACE	1	0.313900	0.3080	1.039	0.3081	1.369
WEXP	1	-0.149796	0.2122	0.498	0.4803	0.861
MAR	1	-0.433704	0.3819	1.290	0.2561	0.648
PARO	1	-0.084871	0.1958	0.188	0.6646	0.919
PRIO	1	0.091497	0.0287	10.200	0.0014	1.096

What are the important predictors of recidivism?

Recidivism Example: (cont'd)

Now, we use the indicators of employment status for each of the 52 weeks in the study, recorded as EMP1-EMP52.

We can fit the model in 2 different ways:

```
proc phreg data=recid;
  model week*arrest(0)=fin age race wexp mar parro prio employed
        / ties=efron;
  array emp(*) emp1-emp52;
  do i=1 to 52;
    if week=i then employed=emp(i);
  end;
run;

*** a shortcut;
proc phreg data=recid;
  model week*arrest(0)=fin age race wexp mar parro prio employed
        / ties=efron;
  array emp(*) emp1-emp52;
  employed=emp(week);
run;
```

The second way takes 23% less time than the first way, but the results are the same.

Recidivism Example: Output

Model WITH employment as time-dependent covariate

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
FIN	1	-0.356722	0.1911	3.484	0.0620	0.700
AGE	1	-0.046342	0.0217	4.545	0.0330	0.955
RACE	1	0.338658	0.3096	1.197	0.2740	1.403
WEXP	1	-0.025553	0.2114	0.015	0.9038	0.975
MAR	1	-0.293747	0.3830	0.488	0.4431	0.745
PARO	1	-0.064206	0.1947	0.109	0.7416	0.938
PRIO	1	0.085139	0.0290	8.644	0.0033	1.089
EMPLOYED	1	-1.328321	0.2507	28.070	0.0001	0.265

Is current employment important?

Do the other covariates change much?

Can you think of any problem with using current employment as a predictor?

Another option for assessing impact of employment

Allison suggests using the employment status of the past week rather than the current week, as follows:

```
proc phreg data=recid;
  where week>1;
  model week*arrest(0)=fin age race wexp mar parro prio employed
        / ties=efron;
  array emp(*) emp1-emp52;
  employed=emp(week-1);
run;
```

The coefficient for EMPLOYED **changes from -1.33 to -0.79, so the risk ratio is about 0.45 instead of 0.27. It is still highly significant with** $\chi^2 = 13.1$**.**

Does this model improve the causal interpretation?

Other options for time-dependent covariates:

- multiple lags of employment status (week-1, week-2, etc.)
- cumulative employment experience (proportion of weeks worked)

Some cautionary notes

- Time-varying covariates must be carefully constructed to ensure interpretability
- There is no point adding a time-varying covariate whose value changes the same as study time ..... you will get the same answer as using a fixed covariate measured at study entry. For example, suppose we want to study the effect of age on time to death.

  We could

  1. use age at start of the study as a fixed covariate
  2. age as a time varying covariate

  However, the results will be the same! Why?

Using time-varying covariates to assess model fit

Suppose we have just fit the following model:

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

E.g., the nursing home data with gender, marital status and health.

Suppose we want to test the proportionality assumption on health (Z_p)

Create a new variable:

$$Z_{p+1}(t) = Z_p * \gamma(t)$$

where $\gamma(t)$ is a known function of time, such as

$$\begin{aligned} \gamma(t) &= t \\ &\text{or } \log(t) \\ &\text{or } e^{-\rho t} \\ &\text{or } I_{\{t>t^*\}} \end{aligned}$$

Then testing $\mathrm{H}_0 : \beta_{p+1} = 0$ is a test for non-proportionality

Illustration: Colon Cancer data

```
*** model without time*covariate interaction;
proc phreg data=surv;
  model survtime*censs(1) = trtm stagen ;
```

Model without time*stage interaction

Event and Censored Values

Total	Event	Censored	Percent Censored
274	218	56	20.44

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	1959.927	1939.654	20.273 with 2 DF (p=0.0001)
Score	.	.	18.762 with 2 DF (p=0.0001)
Wald	.	.	18.017 with 2 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
TRTM	1	0.016675	0.13650	0.01492	0.9028	1.017
STAGEN	1	-0.701408	0.16539	17.98448	0.0001	0.496

```
*** model WITH time*covariate interaction;
proc phreg data=surv ;
  model survtime*censs(1) = trtm stagen tstage ;
  tstage=stagen*exp(-survtime/1000);
```

Model WITH time*stage interaction

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	1959.927	1902.374	57.553 with 3 DF (p=0.0001)
Score	.	.	35.960 with 3 DF (p=0.0001)
Wald	.	.	19.319 with 3 DF (p=0.0002)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
TRTM	1	0.008309	0.13654	0.00370	0.9515	1.008
STAGEN	1	1.402244	0.45524	9.48774	0.0021	4.064
TSTAGE	1	-8.322371	2.04554	16.55310	0.0001	0.000

Like Cox and Oakes, we can run a few different models

Time-varying covariates in Stata

Create a data set with an ID column, and one line per person for each different value of the time varying covariate.

```
. infile id time status group z using cox4_stata.dat
or
. input id time status group    z
      1     3       1      1    0
      2     5       0      1    0
      3     5       1      1    1
      4     6       1      1    0
      5     6       0      1    0
      5     8       0      1    1
      6     4       0      0    1
      7     5       0      0    0
      7     7       1      0    1
      8     8       0      0    0
      9     5       0      0    0
      9     9       1      0    1
     10     3       0      0    0
     10    10       0      0    1
. end
```

```
. stset time status

. cox time group z, dead(status) tvid(id)

------------------------------------------------------------------------------
    time |
  status |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
   group |   1.826757   1.228625      1.487   0.137      -.5813045    4.234819
       z |   .7059632   1.206304      0.585   0.558       -1.65835    3.070276
------------------------------------------------------------------------------
```

Time-varying covariates in Splus

Create a data set with start and stop values of time:

id	start	stop	status	group	z
1	0	3	1	1	0
2	0	5	0	1	0
3	0	5	1	1	1
4	0	6	1	1	0
5	0	6	0	1	0
5	6	8	0	1	1
6	0	4	0	0	1
7	0	5	0	0	0
7	5	7	1	0	1
8	0	8	0	0	0
9	0	5	0	0	0
9	5	9	1	0	1
10	0	3	0	0	0
10	3	10	0	0	1

Then the Splus commands and results are:

Commands:

```
y_read.table("cox4_splus.dat",header=T)

agreg(y$start,y$stop,y$status,cbind(y$group,y$z))
```

Results:

```
 Alive Dead Deleted
     9    5       0

      coef exp(coef) se(coef)     z     p
[1,] 1.827      6.21     1.23 1.487 0.137
[2,] 0.706      2.03     1.21 0.585 0.558

     exp(coef) exp(-coef) lower .95 upper .95
[1,]      6.21      0.161     0.559      69.0
[2,]      2.03      0.494     0.190      21.5

Likelihood ratio test= 2.73  on 2 df,   p=0.256
Efficient score test = 2.73  on 2 df,   p=0.256
```

Piecewise Cox Model: (Collett, Chapter 10)

A time dependent covariate can be used to create a piecewise PH cox model. Suppose we are interested in comparing two treatments, and:

- HR=θ_1 during the interval $(0, t_1)$
- HR=θ_2 during the interval (t_1, t_2)
- HR=θ_3 during the interval (t_2, ∞)

Define the following covariates:

- X - treatment indicator
  ($X = 0 \rightarrow$ standard, $X = 1 \rightarrow$ new treatment)
- Z_2 - indicator of change in HR during 2nd interval

$$Z_2(t) = \begin{cases} 1 & \text{if } t \in (t_1, t_2) \text{ and } X = 1 \\ 0 & \text{otherwise} \end{cases}$$

- Z_3 - indicator of change in HR during 3rd interval

$$Z_3(t) = \begin{cases} 1 & \text{if } t \in (t_2, \infty) \text{ and } X = 1 \\ 0 & \text{otherwise} \end{cases}$$

The model for the hazard for individual i is:

$$\lambda_i(t) = \lambda_0(t) \exp\{\beta_1 x_i + \beta_2 z_{2i}(t) + \beta_3 z_{3i}(t)\}$$

What are the log hazard ratios for an individual on the new treatment relative to one on the standard treatment?

Time varying (or time-dependent) covariates

Case Study of MAC Disease Trial

ACTG 196 was a randomized clinical trial to study the effects of combination regimens on prevention of MAC (mycobacterium avium complex) disease, which is one of the most common opportunistic infections in AIDS patients and is associated with high mortality and morbidity.

The **treatment regimens** were:

- clarithromycin (new)
- rifabutin (standard)
- clarithromycin plus rifabutin

This trial enrolled patients between April 1993 and February 1994, and followed patients through August 1995. In February of 1994, the dosage of rifabutin was reduced from 3 capsules per day (450mg) to 2 capsules per day (300mg) due to concern over **uveitis**, an adverse experience resulting in inflammation of the uveal tract in the eyes (about 3-4% of patients reported uveitis). All patients were to reduce their dosage by March 8, 1994. However, some patients had already discontinued the treatment, died, or discontinued the study.

The main intent-to-treat analysis compared the 3 treatment arms without adjusting for this change in dosage.

Other supporting analyses attempted to untangle the effect of this **"study wide dose reduction" (SWDR)**.

Proportion on each treatment arm with SWDR

Treatment by study wide dose reduction

TABLE OF TRTMT BY SWDRSTAT

TRTMT	SWDRSTAT(Study Wide Dose Reduction Status)		
Frequency|			
Row Pct |	No	Yes	Total
R	125	266	391
	31.97	68.03	
C+R	170	219	389
	43.70	56.30	
C	124	274	398
	31.16	68.84	
Total	419	759	1178

STATISTICS FOR TABLE OF TRTMT BY SWDRSTAT

Statistic	DF	Value	Prob
Chi-Square	2	16.820	0.001
Likelihood Ratio Chi-Square	2	16.610	0.001
Mantel-Haenszel Chi-Square	1	0.067	0.795
Phi Coefficient		0.119	
Contingency Coefficient		0.119	
Cramer's V		0.119	

Sample Size = 1178

Original Logrank test Comparing 3 Treatment Arms (How would you get pairwise tests?)

Dependent Variable: MACTIME Time to MAC disease (days)
Censoring Variable: MACSTAT MAC status (1=yes,0=censored)
Censoring Value(s): 0
Ties Handling: BRESLOW

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
1178	121	1057	89.73

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	1541.064	1525.932	15.133 with 2 DF (p=0.0005)
Score	.	.	15.890 with 2 DF (p=0.0004)
Wald	.	.	15.209 with 2 DF (p=0.0005)

Analysis of Maximum Likelihood Estimates

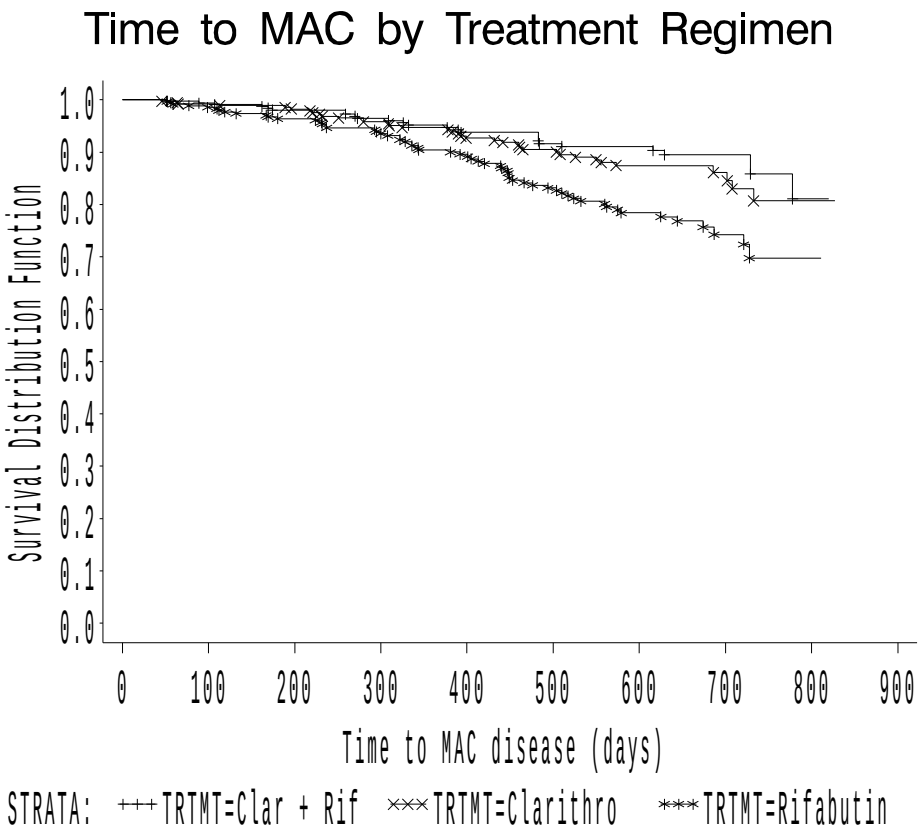
Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
CLARI	1	0.231842	0.25748	0.81074	0.3679	1.261
RIF	1	0.826883	0.23601	12.27480	0.0005	2.286

Variable	Label
CLARI	1=Clarithromycin arm, 0 otherwise
RIF	1=Rifabutin arm, 0 otherwise

Linear Hypotheses Testing

Label	Wald Chi-Square	DF	Pr > Chi-Square
TEST_TRT	15.2094	2	0.0005

Kaplan-Meier Survival Plot
Estimated Probabilities of Remaining MAC-free

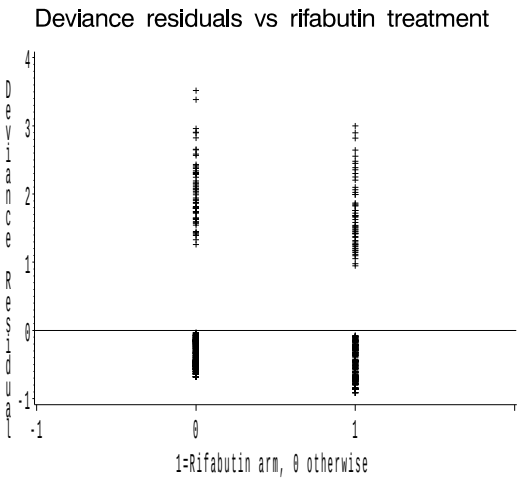


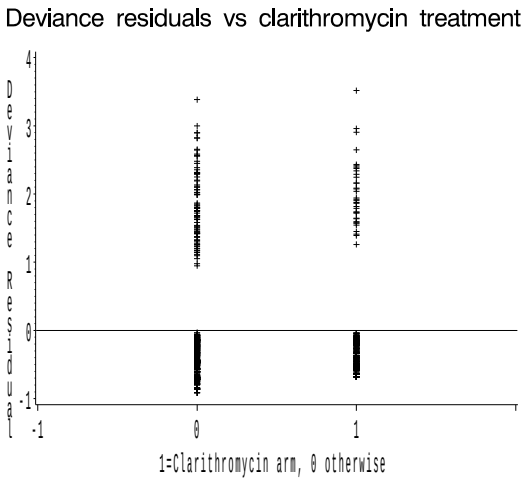

```
%ps(mactrt.ps,mode=replace);
proc lifetest data=weighted noprint outsurv=survres
     graphics nocens plots=(s);
  time mactime*macstat(0);
  strata trtmt;
  title 'Time to MAC by Treatment Regimen';
  format trtmt trtfmt.;
run;
```

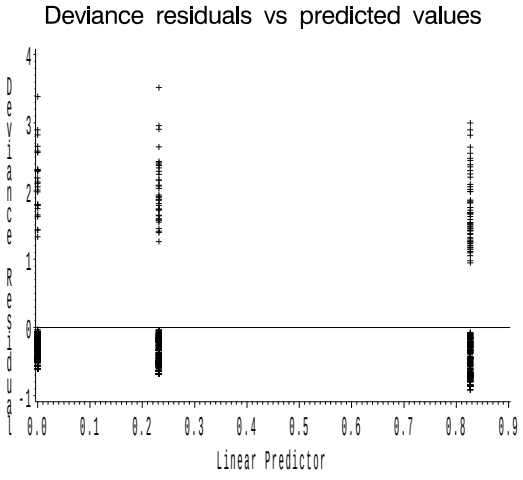
How well does this model fit?

Let's take a look at the residual plots...

First, the deviance residuals:



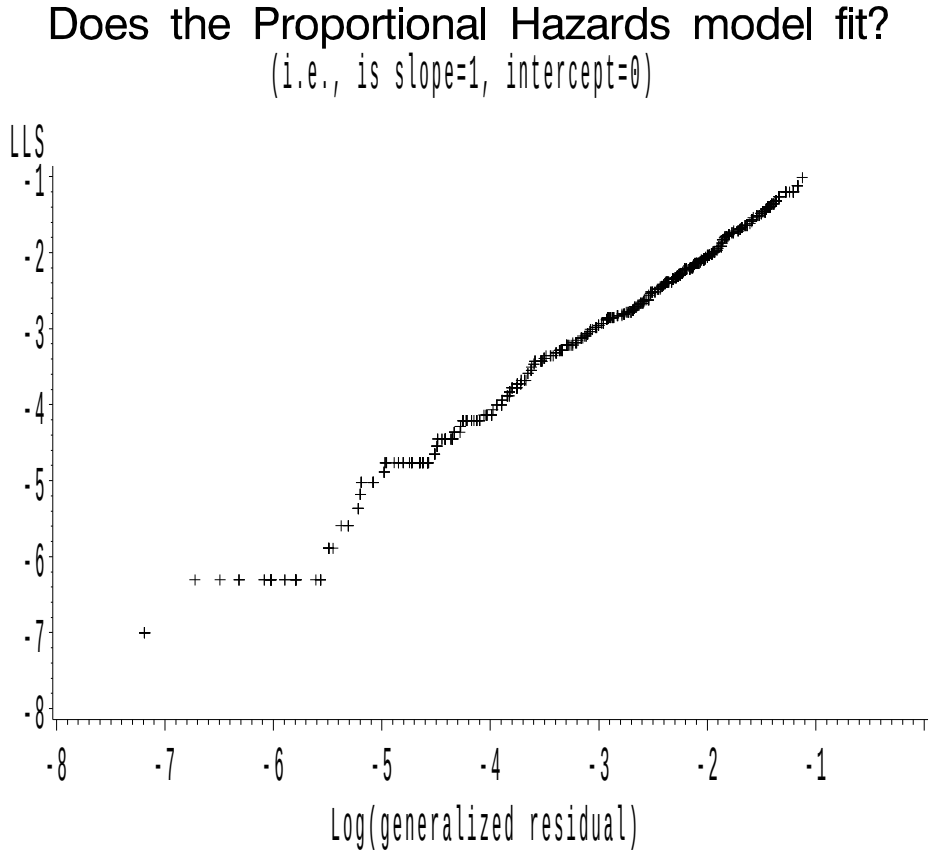




Plotting deviance residuals vs binary covariates is not very useful.

How about the generalized residuals?

(Are they like a sample from a censored unit exponential?)

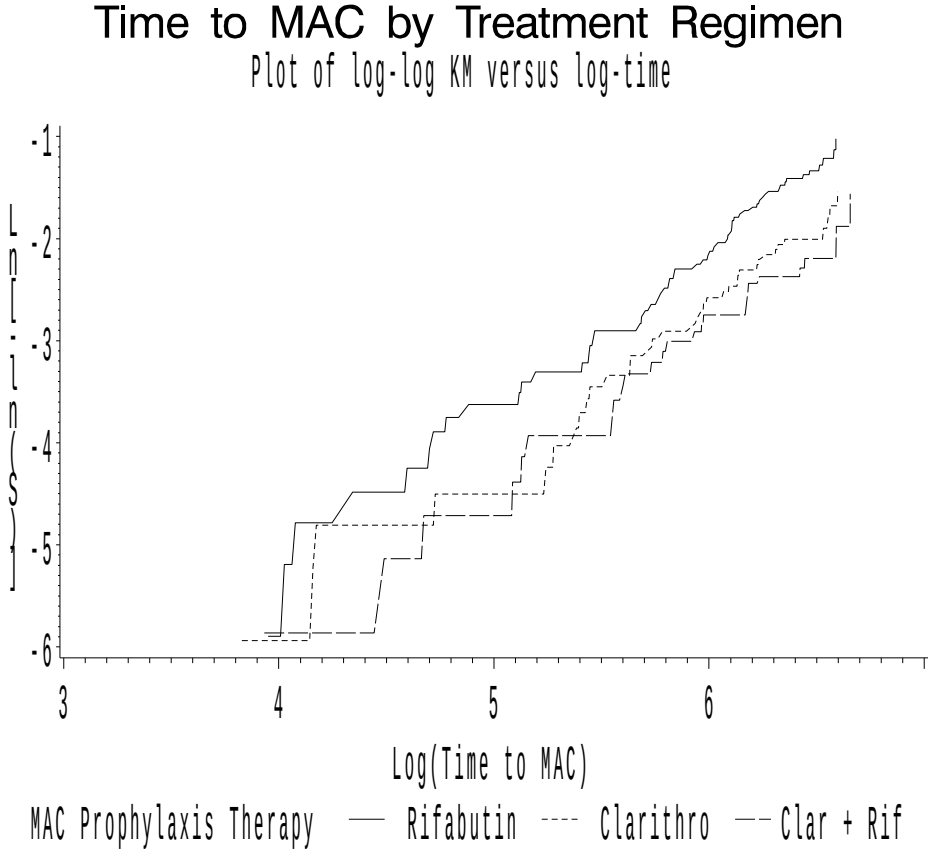


intercept=0.056

slope=1.028

(based on fitting a regression line to residuals)

We can also look at the log cumulative hazard plots (i.e., $\log[-\log(\hat{S})]$**) versus log time to see whether the lines are parallel for the three treatment groups.**



(I have joined the individual points using i=join in the symbol statement, to make them easier to see.)

Shouldn't we adjust for Baseline CD4 count?

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	1541.064	1488.737	52.328 with 3 DF (p=0.0001)
Score	.	.	43.477 with 3 DF (p=0.0001)
Wald	.	.	43.680 with 3 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
CLARI	1	0.198798	0.25747	0.59619	0.4400	1.220
RIF	1	0.837240	0.23598	12.58738	0.0004	2.310
CD4	1	-0.019641	0.00367	28.59491	0.0001	0.981

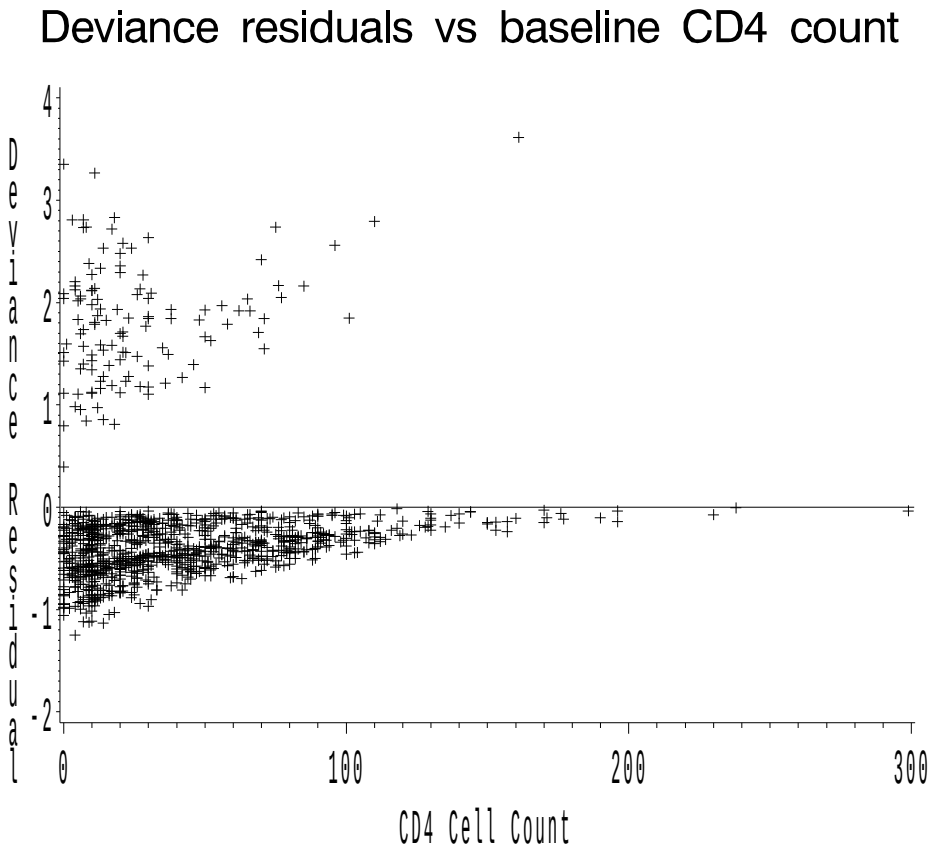
Analysis of Maximum Likelihood Estimates

Variable	Label
CLARI	1=Clarithromycin arm, 0 otherwise
RIF	1=Rifabutin arm, 0 otherwise
CD4	CD4 Cell Count

Is CD4 count a confounder?

(An analysis stratified by CD4 category gave almost identical results. Other important covariates included CTG (clinical trials group) and Karnofsky status).

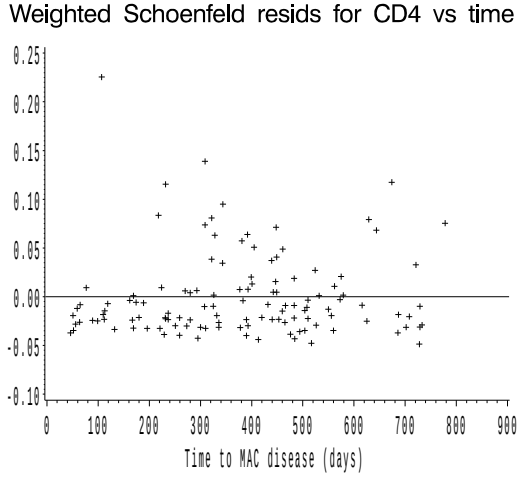
What do the deviance residuals look like versus a continuous covariate, like CD4?



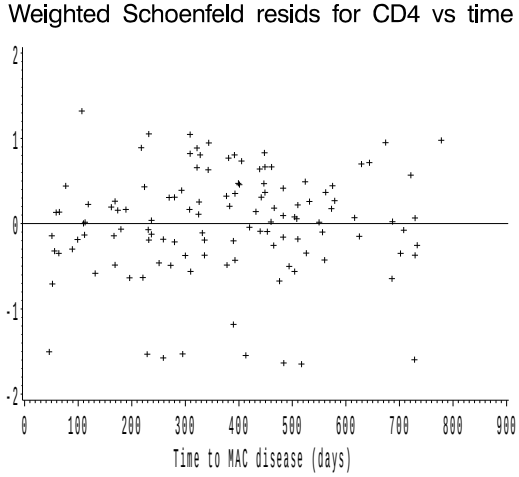
We might want to consider some kind of transformation of CD4 count (like log or square root). If we don't feel comfortable with the linearity of CD4 count, we can also dichotomize it (CD4CAT).

Another way of checking the proportionality assumption is by using the Weighted Schoenfeld residual plots for each covariate

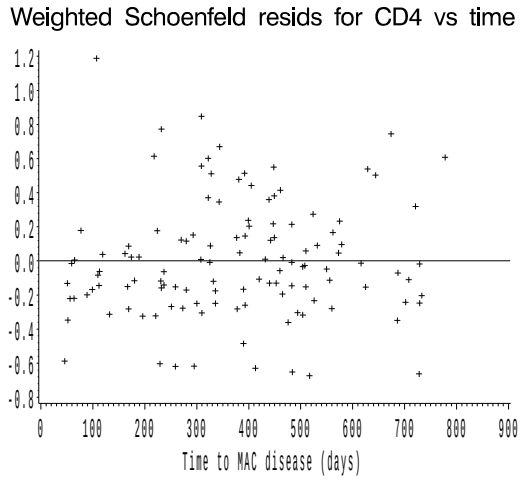
Raw CD4 count



log CD4 count



Square root CD4 count



So far, the graphical techniques have not indicated any major departure from proportional hazards. However, we can test this formally by creating a time dependent covariate for rifabutin and clarithromycin:

```
riftd=rif*((mactime-365)/30);

claritd=clari*((mactime-365)/30);
```

Even though the dose reduction was only for rifabutin, patients on all 3 arms had to have the dose reduction ... they just took 2 capsules of their placebo, and didn't know whether it was placebo or active drug.

I have centered the time-dependent covariates at 365 days (one year), so that the HR for rif alone and clari alone will apply at one year. Then I have divided by 30, so that the resulting HR can be interpreted as the change for each month away from 365 days.

Question: Can we do this within a data step using the above statements, or do these statements need to be given in the PROC PHREG procedure?

Time-dependent covariates for clari and rif

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	1541.064	1525.837	15.227 with 4 DF (p=0.0043)
Score	.	.	16.033 with 4 DF (p=0.0030)
Wald	.	.	15.327 with 4 DF (p=0.0041)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
CLARI	1	0.229811	0.25809	0.79287	0.3732	1.258
RIF	1	0.823227	0.23624	12.14274	0.0005	2.278
CLARITD	1	0.003065	0.04073	0.00566	0.9400	1.003
RIFTD	1	0.010627	0.03765	0.07965	0.7778	1.011

Analysis of Maximum Likelihood Estimates

Variable	Label
CLARI	1=Clarithromycin arm, 0 otherwise
RIF	1=Rifabutin arm, 0 otherwise

Neither time-dependent covariate was significant.

This analysis also indicated that there are no major departures from proportional hazards for the three treatment arms.

However, it may still be the case that having the study-wide dose reduction had some relationship with MAC disease.

We can assess this by creating a time dependent variable for the SWDR.

We'll look at the following models:

(1) SWDRSTAT as a simple indicator

(2) SWDRSTAT and SWDRTD, with

```
swdrtd = swdrstat*((mactime-365)/30)
```

(3) SWDR as time dependent covariate

Naive model with fixed SWDR indicator (SWDRSTAT):

```
 Dependent Variable: MACTIME   Time to MAC disease (days)
 Censoring Variable: MACSTAT   MAC status (1=yes,0=censored)
 Censoring Value(s): 0
 Ties Handling: BRESLOW
```

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	1541.064	1495.857	45.208 with 3 DF (p=0.0001)
Score	.	.	51.497 with 3 DF (p=0.0001)
Wald	.	.	48.749 with 3 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
CLARI	1	0.449936	0.26142	2.96236	0.0852	1.568
RIF	1	1.006639	0.23852	17.81114	0.0001	2.736
SWDRSTAT	1	-1.125032	0.19283	34.04055	0.0001	0.325

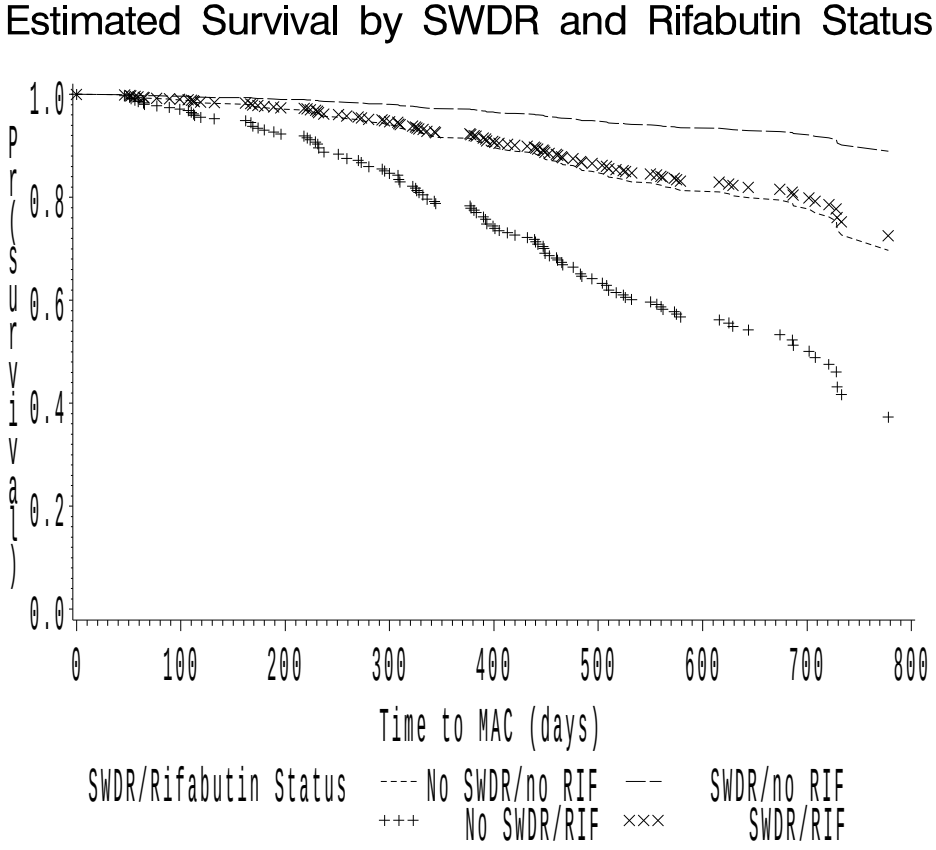
Analysis of Maximum Likelihood Estimates

Variable	Label
CLARI	1=Clarithromycin arm, 0 otherwise
RIF	1=Rifabutin arm, 0 otherwise
SWDRSTAT	Study Wide Dose Reduction Status

Reduction of dosage from 450mg to 300mg appears to be protective, which seems counter-intuitive

Predicted Baseline Survival Curves:

Another way to see this is through the predicted baseline survival curves. The two lines are for those <u>not</u> on rifabutin, while the x's and +'s are for those on rifabutin. In each case, the higher line (better prognosis) of the pair is for those who did have the SWDR.



Test for proportionality:

```
proc phreg data=weighted;
  model mactime*macstat(0) = clari rif swdrstat swdrtd;

  *** create time by covariate interaction for swdr status;
  swdrtd=swdrstat*((mactime-365)/30);

  test_trt: test clari, rif;
  title 'Test of treatment Differences';
  title2 'and test of proportionality at t=365 days';
```

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	1541.064	1492.692	48.372 with 4 DF (p=0.0001)
Score	.	.	55.174 with 4 DF (p=0.0001)
Wald	.	.	50.719 with 4 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
CLARI	1	0.430051	0.26126	2.70947	0.0998	1.537
RIF	1	1.005416	0.23845	17.77884	0.0001	2.733
SWDRSTAT	1	-1.126498	0.19752	32.52551	0.0001	0.324
SWDRTD	1	0.055550	0.03201	3.01112	0.0827	1.057

Variable	Label
CLARI	1=Clarithromycin arm, 0 otherwise
RIF	1=Rifabutin arm, 0 otherwise
SWDRSTAT	Study Wide Dose Reduction Status
SWDRTD	swdrstat*((mactime-365)/30)

Interpretation of Hazard Ratios

$$\beta_{\text{swdrstat}} = -1.1265$$

$$\beta_{\text{swdrtd}} = 0.0556$$

Time (months)	Time (days)	calculation	Hazard Ratio
6	182.5	$\exp[-1.1265 + (-6.08)(0.0556)]$	0.231
12	**365**	$\exp[-1.1265 + (0)(0.0556)]$	0.324
18	547.5	$\exp[-1.1265 + (6.08)(0.0556)]$	0.454
24	730	$\exp[-1.1265 + (12.17)(0.0556)]$	0.637
30	912.5	$\exp[-1.1265 + (18.25)(0.0556)]$	0.893
36	1095	$\exp[-1.1265 + (24.33)(0.0556)]$	1.253

$$HR = \exp[\beta_{\text{swdrstat}} + \beta_{\text{swdrtd}} \left(\frac{\text{mactime}-365)}{30}\right)]$$

In the early period after randomization to treatment, reduction of randomized dosage from 450mg to 300mg is associated with a decreased risk of MAC disease. After taking the higher dosage for about 32 months, dropping to the lower dosage has no impact, and as the treatment time increases beyond 32 months, a lower dosage tends to be associated with increased risk of MAC.

3 different ways to code SWDR as time-dependent covariate

```
proc phreg data=weighted;
  model mactime*macstat(0) = clari rif swdr;

  if (swdrtime>=mactime) then swdr=0;
  else do;
    if swdrstat=1 then swdr=1;
    else swdr=0;
  end;

  test_trt: test clari, rif;
  title2 'I. Time-dependent indicator of dose reduction';



proc phreg data=weighted;
  model mactime*macstat(0) = clari rif swdr;

  if swdrstat=0 or (swdrtime>=mactime) then swdr=0;
  else swdr=1;

  test_trt: test clari, rif;
  title2 'II. Time-dependent indicator of dose reduction';



proc phreg data=weighted;
  model mactime*macstat(0) = clari rif swdr;

  if swdrstat=1 and (swdrtime<mactime) then swdr=1;
  else swdr=0;

  test_trt: test clari, rif;
  title2 'III. Time-dependent indicator of dose reduction';
```

Output is the same for all 3 cases:

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
1178	121	1057	89.73

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	1541.064	1517.426	23.639 with 3 DF (p=0.0001)
Score	.	.	24.844 with 3 DF (p=0.0001)
Wald	.	.	24.142 with 3 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
CLARI	1	0.328849	0.26017	1.59762	0.2062	1.389
RIF	1	0.905299	0.23775	14.49956	0.0001	2.473
SWDR	1	-0.648887	0.21518	9.09389	0.0026	0.523

SWDR is still protective? Does this make sense intuitively?

What other methods can we use to account for change in dosage?

Weighted adjusted dose (WAD) analyses

To try to get a better idea of the effect of changing doses of rifabutin on the hazard for MAC disease, I created the following weighted dose of randomized rifabutin:

- Between randomization date and SWDR date
  $\Longrightarrow$ # Days at 450mg
- Between SWDR date and off-study date
  $\Longrightarrow$ # Days at 300mg
- Between randomization date and Off-study date
  $\Longrightarrow$ # Total Days
- Weighted randomized dose

  ```
  rifwadr = (days450 + days300)/totdays
  ```

- Transformed to number of capsules per day;

  ```
  rifwadr=rifwadr/150;
  ```

- Also calculated weighted dose *while on treatment* by starting with on treatment date, stopping with off-treatment date, and dividing by the total days on study.

Weighted adjusted dose (WAD) analyses
Randomized assignment to rifabutin

```
Dependent Variable: MACTIME   Time to MAC disease (days)
Censoring Variable: MACSTAT   MAC status (1=yes,0=censored)
Censoring Value(s): 0
Ties Handling: BRESLOW
```

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
1178	121	1057	89.73

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	1541.064	1493.476	47.588 with 3 DF (p=0.0001)
Score	.	.	52.770 with 3 DF (p=0.0001)
Wald	.	.	50.295 with 3 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
CLARI	1	0.453283	0.26119	3.01179	0.0827	1.573
RIF	1	1.004846	0.23826	17.78681	0.0001	2.731
RIFWADR	1	1.530462	0.25681	35.51502	0.0001	4.620

For each additional capsule of rifabutin specified as randomized treatment, the HR for MAC increased by 4.6 times

Weighted adjusted dose (WAD) analyses
Actual dosage of rifabutin during the study

```
Dependent Variable: MACTIME   Time to MAC disease (days)
Censoring Variable: MACSTAT   MAC status (1=yes,0=censored)
Censoring Value(s): 0
Ties Handling: BRESLOW
```

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	1541.064	1489.993	51.071 with 3 DF (p=0.0001)
Score	.	.	55.942 with 3 DF (p=0.0001)
Wald	.	.	53.477 with 3 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
CLARI	1	0.489583	0.26256	3.47693	0.0622	1.632
RIF	1	1.019675	0.23873	18.24291	0.0001	2.772
RIFWAD	1	-0.664689	0.10686	38.69332	0.0001	0.514

Here, higher values of RIFWAD probably reflect that the patient was able to stay on treatment longer, which was protective. The SWDR variable is also capturing whether a patient had been able to tolerate the treatment long enough to have the chance to have the protocol-mandated dose reduction.

What happens if we add treatment discontinuation as a time dependent covariate?

```
Dependent Variable: MACTIME   Time to MAC disease (days)
Censoring Variable: MACSTAT   MAC status (1=yes,0=censored)
Censoring Value(s): 0
Ties Handling: BRESLOW
```

Summary of the Number of Event and Censored Values

Total	Event	Censored	Percent Censored
1178	121	1057	89.73

Testing Global Null Hypothesis: BETA=0

Criterion	Without Covariates	With Covariates	Model Chi-Square
-2 LOG L	1541.064	1501.595	39.469 with 4 DF (p=0.0001)
Score	.	.	42.817 with 4 DF (p=0.0001)
Wald	.	.	41.027 with 4 DF (p=0.0001)

Analysis of Maximum Likelihood Estimates

Variable	DF	Parameter Estimate	Standard Error	Wald Chi-Square	Pr > Chi-Square	Risk Ratio
CLARI	1	0.420447	0.26111	2.59284	0.1073	1.523
RIF	1	0.984114	0.23847	17.02975	0.0001	2.675
SWDR	1	-0.139245	0.23909	0.33919	0.5603	0.870
RXSTOP	1	0.902592	0.21792	17.15473	0.0001	2.466

SWDR is no longer significant!

Last of all, a comparison of some of these models:

Model terms	q	$-2 \log L$	AIC Criterion
Clari, Rif	2	1525.93	1531.93
Clari, Rif, Cd4cat	3	1497.57	1506.57
Clari, Rif, Cd4	3	1488.74	1497.74
Clari, Rif, Cd4cat, Ctg, Karnof	5	1482.67	1497.67
Clari, Rif, Swdrstat	3	1495.86	1504.86
Clari, Rif, Rifwadr	3	1493.48	1502.48
Clari, Rif, Swdrstat, Rifwadr	4	1493.44	1505.44
Clari, Rif, Rifwad	3	1489.99	1498.99
Models with time-dependent covariates			
Clari, Rif, Claritd, Riftd	4	1525.84	1537.84
Clari, Rif, Swdrstat, Swdrtd	4	1492.69	1504.69
Clari, Rif, Swdr	3	1517.43	1526.43
Clari, Rif, Swdr, Rxstop	4	1501.60	1513.60
Clari, Rif, Cd4cat, Karnof, Rxstop	5	1461.90	1476.90
Clari, Rif, Cd4cat, Karnof, Rifwad	5	1448.14	1463.14

Parametric Survival Analysis

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

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

We used the following estimating approach:

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

Both estimates were based on maximum likelihood theory.

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

- The assumption of proportional hazards might not be appropriate (based on major departures)
- If a parametric model actually holds, then we would probably gain efficiency
- We may want to handle non-standard situations like
  - interval censoring
  - incorporating population mortality
- We may want to make some connections with other familiar approaches (e.g. use of the Poisson likelihood)
- We may want to obtain some estimates for use in designing a future survival study.

A simple start: Exponential Regression

- **Observed data:** $(X_i, \delta_i, \mathbf{Z}_i)$ for individual i, $\mathbf{Z}_i = (Z_{i1}, Z_{i2}, ..., Z_{ip})$ represents a set of p covariates.

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

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

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

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

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

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

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

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

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

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

WHY?

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

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

Exponential regression for the 2-sample case:

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

  Hazard Rate:

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

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

- **What is the hazard for group 0?**

- **What is the hazard for group 1?**

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

- **What is the interpretation of β_1?**

Likelihood for Exponential Model

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

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

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

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

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

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

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

Maximum Likelihood for Exponential

How do we use the likelihood?

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

The log-likelihood is:

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

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

$$\log \mathcal{L} = \sum_i \left[\delta_i \log(\Psi(\mathbf{Z}_i)) \right] - \sum_i \Psi(\mathbf{Z}_i) X_i \qquad (1)$$

General Form of Log-likelihood for Right Censored Data

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

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

where

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

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

Consider our model for the hazard rate:

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

We can write this using vector notation, as follows:

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

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

Then, we can write the hazard as:

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

Now we can substitute $\Psi(\mathbf{Z}_i) = \exp[\boldsymbol{\beta}\mathbf{Z}_i]$ in the log-likelihood shown in (1):

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

Score Equations

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

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

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

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

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

To find the variance of the MLE's, we need to take the second derivatives:

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

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

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

where

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

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

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

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

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

First, what is the MLE of β_0?

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

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

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

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

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

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

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

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

By the delta method,

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

The Two-Sample Problem:

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

The log-likelihood:

$$\log \mathcal{L} = \sum_{i=1}^{n} \delta_i(\beta_0 + \beta_1 Z_i) - \sum_{i=1}^{n} X_i \exp(\beta_0 + \beta_1 Z_i)$$

so
$$\frac{\partial \log \mathcal{L}}{\partial \beta_0} = \sum_{i=1}^{n} [\delta_i - X_i \exp(\beta_0 + \beta_1 Z_i)]$$
$$= (d_0 + d_1) - (t_0 e^{\beta_0} + t_1 e^{\beta_0 + \beta_1})$$

$$\frac{\partial \log \mathcal{L}}{\partial \beta_1} = \sum_{i=1}^{n} Z_i [\delta_i - X_i \exp(\beta_0 + \beta_1 Z_i)]$$
$$= d_1 - t_1 e^{\beta_0 + \beta_1}$$

This implies:
$$\hat{\lambda}_1 = e^{\hat{\beta}_0 + \hat{\beta}_1} = ?$$
$$\hat{\lambda}_0 = e^{\hat{\beta}_0} = ?$$
$$\hat{\beta}_0 = ?$$
$$\hat{\beta}_1 = ?$$

Important Result:

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

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

Exponential Regression: Means and Medians

Mean Survival Time

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

- **Control Group:**

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

- **Treatment Group:**

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

Median Survival Time

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

- **Control Group:**

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

- **Treatment Group:**

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

Exponential Regression:
Variance Estimates and Test Statistics

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

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

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

So our test statistics are formed as:

For testing $H_o : \beta_0 = 0$:

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

For testing $H_o : \beta_1 = 0$:

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

How would we form confidence intervals for the hazard ratio?

The Likelihood Ratio Test Statistic:

(An alternative to the Wald test)

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

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

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

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

where d is the number of deaths or failures.

The log-likelihood is

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

and the MLE is

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

2-Sample Case: LR test calculations

Data:

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

Under the **alternative hypothesis**:

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

The MLE's are:

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

Under the **null hypothesis**:

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

The corresponding MLE is

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

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

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

Nursing home example:
<u>For the females:</u>

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

<u>For the males:</u>

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

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

Hand Calculations using events and follow-up:

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

- $d_0 = 902$ (females)
  $d_1 = 367$ (males)

- $t_0 = 310754$ (female follow-up)
  $t_1 = 75457$ (male follow-up)

- This yields an estimated log HR:

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

- The estimated standard error is:

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

- So the Wald test becomes:

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

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

Exponential Regression in STATA

```
. use nurshome

. stset los fail

. streg gender, dist(exp) nohr

        failure _d:  fail
  analysis time _t:  los


Iteration 0:   log likelihood = -3352.5765
Iteration 1:   log likelihood =  -3321.966
Iteration 2:   log likelihood = -3320.4792
Iteration 3:   log likelihood = -3320.4766
Iteration 4:   log likelihood = -3320.4766

Exponential regression -- log relative-hazard form

No. of subjects =         1591             Number of obs   =      1591
No. of failures =         1269
Time at risk    =       386211
                                           LR chi2(1)      =     64.20
Log likelihood  =   -3320.4766             Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gender |    .516186   .0619148     8.337   0.000     .3948352    .6375368
       _cons |  -5.842142   .0332964  -175.459   0.000    -5.907402   -5.776883
------------------------------------------------------------------------------
```

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

Exponential Regression in SAS - PROC LIFEREG

```
proc format;
  value censfmt    1='Censored'
                   0='Dead';

  value grpfmt     0='Group 0 (F)'
                   1='Group 1 (M)';

Title 'Exponential Hazard Model for Nursing Home Patients';
data morris;
  infile 'ch12.dat';
  input los age trt gender marstat hltstat cens;

data morris2;
  set morris;
  if los=0 then delete;

proc freq data=morris2;
  table cens*gender/ norow nocol nopercent;
  format cens censfmt. gender grpfmt.;

proc lifereg data=pop covout outest=survres;
  model los*censor(1)=gender /dist=exponential;
run;
```

RESULTS:

TABLE OF CENS BY GENDER

CENS	GENDER		
Frequency	Group 0 (F)	Group 1 (M)	Total
Event	902	367	1269
Censored	271	51	322
Total	1173	418	1591

PROC LIFEREG RESULTS:

```
          Exponential Hazard Model for Nursing Home Patients

                          Lifereg  Procedure

Data Set          =WORK.MORRIS2
Dependent Variable=Log(LOS)
Censoring Variable=CENS
Censoring Value(s)=     1
Noncensored Values=  1269  Right Censored Values=    322
Left Censored Values=   0  Interval Censored Values=   0


Log Likelihood for EXPONENT -3320.476626

                          Lifereg  Procedure

Variable  DF   Estimate  Std Err ChiSquare  Pr>Chi Label/Value

INTERCPT   1 5.84213388 0.033296     30786  0.0001 Intercept
GENDER     1 -0.5161878 0.061915  69.50734  0.0001
SCALE      0          1        0                   Extreme value scale
```

Note that the estimates for β_0 and β_1 above are the opposites of what we calculated. I'll explain why the output has this form when we get to AFT models.

The Weibull Regression Model

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

$$S(t) \;=\; \exp[-\lambda(t^{\kappa})]$$

and the hazard function is:

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

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

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

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

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

For the 2-sample problem, we have:

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

Weibull MLEs for the 2-sample problem:

Log-likelihood:

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

$$\begin{aligned}
\Rightarrow \qquad \exp(\hat{\beta}_0) &= d_0/t_{0\kappa} \\
\exp(\hat{\beta}_0 + \hat{\beta}_1) &= d_1/t_{1\kappa} \\
\text{where} \quad t_{j\kappa} &= \sum_{i=1}^{n_j} X_i^{\hat{\kappa}} \;\; \text{among } n_j \text{ subjects} \\
\hat{\lambda}_0(t) &= \hat{\kappa}\; \exp(\hat{\beta}_0)\; t^{\hat{\kappa}-1} \\
\hat{\lambda}_1(t) &= \hat{\kappa}\; \exp(\hat{\beta}_0 + \hat{\beta}_1)\; t^{\hat{\kappa}-1} \\
\widehat{HR} &= \hat{\lambda}_1(t)/\hat{\lambda}_0(t) = \exp(\hat{\beta}_1) \\
&= \exp\left(\frac{d_1/t_{1\kappa}}{d_0/t_{0\kappa}}\right)
\end{aligned}$$

Weibull Regression:
Means and Medians

Mean Survival Time

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

- **Control Group:**

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

- **Treatment Group:**

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

Median Survival Time

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

- **Control Group:**

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

- **Treatment Group:**

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

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

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

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

In cases where x is not an integer, this function has to be evaluated numerically.

The Weibull regression model is very easy to fit:

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

Note: to get more information on these modeling procedures, use the online help facilities. For example, in STATA, you can type:

```
.help streg
```

Weibull in Stata:

```
. streg gender, dist(weibull) nohr

        failure _d:  fail
  analysis time _t:  los


Fitting constant-only model:

Iteration 0:   log likelihood = -3352.5765
Iteration 1:   log likelihood =  -3074.978
Iteration 2:   log likelihood = -3066.1526
Iteration 3:   log likelihood =  -3066.143
Iteration 4:   log likelihood =  -3066.143

Fitting full model:

Iteration 0:   log likelihood =  -3066.143
Iteration 1:   log likelihood = -3045.8152
Iteration 2:   log likelihood = -3045.2772
Iteration 3:   log likelihood = -3045.2768
Iteration 4:   log likelihood = -3045.2768

Weibull regression -- log relative-hazard form

No. of subjects =         1591                 Number of obs   =      1591
No. of failures =         1269
Time at risk    =       386211
                                               LR chi2(1)      =     41.73
Log likelihood  =   -3045.2768                 Prob > chi2     =    0.0000

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gender |   .4138082   .0621021     6.663   0.000    .2920903    .5355261
       _cons |  -3.536982   .0891809   -39.661   0.000   -3.711773   -3.362191
-------------+----------------------------------------------------------------
       /ln_p |  -.4870456   .0232089   -20.985   0.00    -.5325343   -.4415569
-------------+----------------------------------------------------------------
           p |    .614439   .0142605                      .5871152    .6430345
         1/p |   1.627501   .0377726                      1.555127    1.703243
------------------------------------------------------------------------------
```

Weibull in SAS

```
proc lifereg data=morris2 covout outest=survres;
  model los*censor(1)=gender / dist=weibull;
run;
```

```
Data Set          =WORK.MORRIS2
Dependent Variable=Log(LOS)
Censoring Variable=CENS
Censoring Value(s)=     1
Noncensored Values=  1269         Right Censored Values=    322
Left Censored Values=   0         Interval Censored Values=   0


Log Likelihood for WEIBULL -3045.276811


                        Lifereg  Procedure

Variable  DF   Estimate  Std Err ChiSquare  Pr>Chi Label/Value

INTERCPT   1 5.75644118   0.0542  11280.04  0.0001 Intercept
GENDER     1 -0.6734732 0.101067  44.40415  0.0001
SCALE      1 1.62750085 0.037773                  Extreme value scale
```

In SAS, both the exponential and Weibull are special cases of the general class of **accelerated life models** and the parameter interpretations follow from this approach.

To translate the output of SAS (or Stata using the `ereg` command) for Weibull regression, we have to take the negative of the numbers in the output, divided by the "_SCALE_" parameter (σ, or $1/\kappa$).

- $\hat{\beta}_0 = -\text{INTERCPT}/\_scale\_$
- $\hat{\beta}_1 = -\text{COVARIATE}/\_scale\_$

Then we calculate the estimated HR as $\exp(\hat{\beta}_1)$.

The MLE's are:

- $\hat{\beta}_0 = -\text{INTERCPT}/\_scale\_ = -5.756/1.627 = -3.537$
- $\hat{\beta}_1 = -\text{COVARIATE}/\_scale\_ = 0.6735/1.625 = 0.414$

and the estimated HR is $\widehat{HR} = \exp(\hat{\beta}_1) = \exp(0.414) = 1.513$.

Weibull Regression:
Variance Estimates and Test Statistics

It is not so easy to get variance estimates from the output of PROC LIFEREG in SAS or *weibull* in STATA, at least for the parameters we're interested in.

The variances depend on inverting a (3×3) matrix corresponding to the parameters β_0, β_1, and κ. The MLE for $\hat{\kappa}$ has to be obtained numerically (i.e., no closed form), so the standard errors also have to be obtained by computer.

Main objective: to obtain $s.e.(\hat{\beta}_1)$, so that we can form tests and confidence intervals for the hazard ratio.

The output gives us $\hat{\beta}_1^*$ and $s.e.(\hat{\beta}_1^*)$, where $\hat{\beta}_1 = -\hat{\beta}_1^*/\hat{\sigma}$. **If** σ was a constant, then we could just compute

$$var(\hat{\beta}_1) \;=\; \frac{1}{\hat{\sigma}^2}\, var(\hat{\beta}_1^*)$$

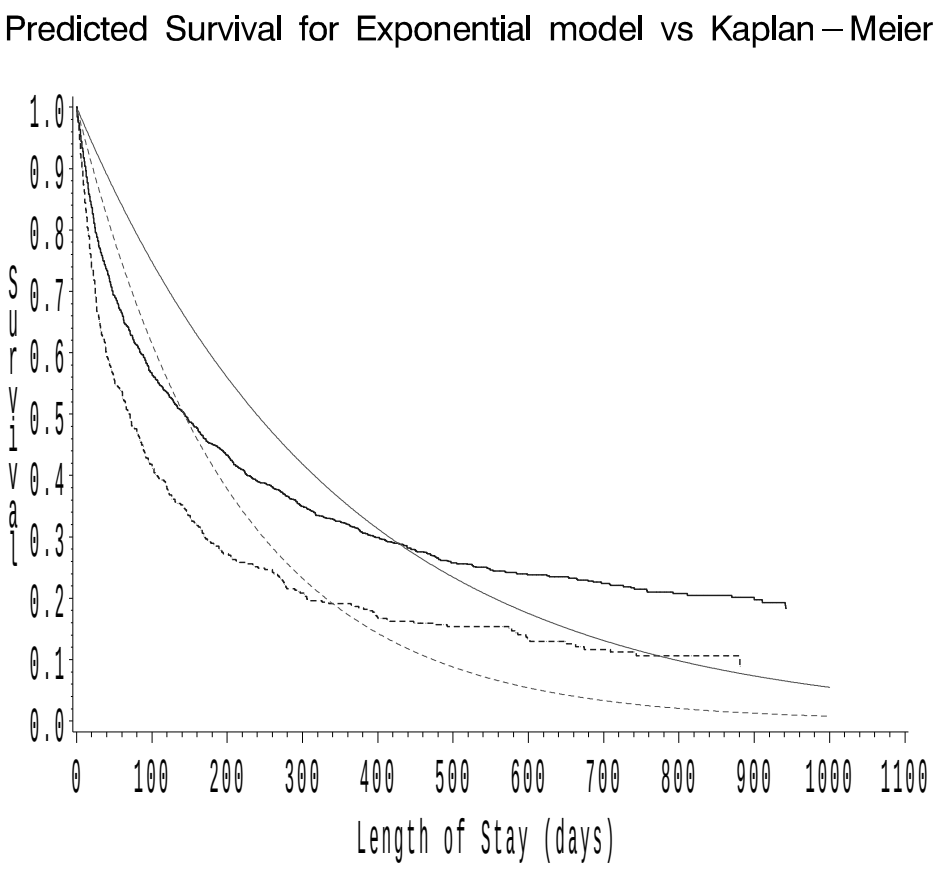
but σ is **also** a random variable! Instead, you need to use an approximation for the variance of a ratio of two random variables:

$$var(\hat{\beta}_1) \;=\; \frac{1}{\hat{\sigma}^4}\left[\hat{\sigma}^2 var(\hat{\beta}_1^*) + (\hat{\beta}_1^*)^2 var(\hat{\sigma}) - 2\hat{\beta}_1^*\hat{\sigma}\, cov(\hat{\beta}_1^*, \hat{\sigma})\right]$$

where you get $var(\hat{\beta}_1^*)$ and $var(\hat{\sigma})$ by squaring the standard errors of the COVARIATE term and SCALE term, respectively, from the PROC LIFEREG or *weibull* output.

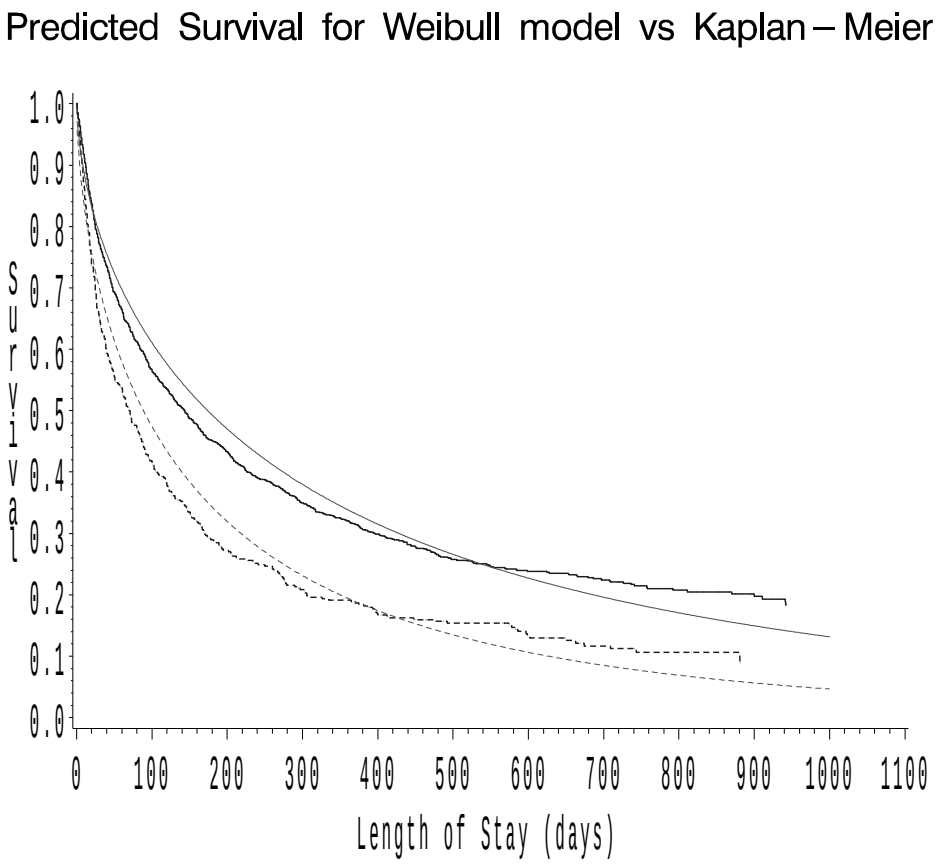
Comparison of Exponential with Kaplan-Meier

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



Comparison of Weibull with Kaplan-Meier

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



Which do you think fits best?

Other useful plots for evaluating fit to exponential and Weibull models

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

Why are these useful?

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

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

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

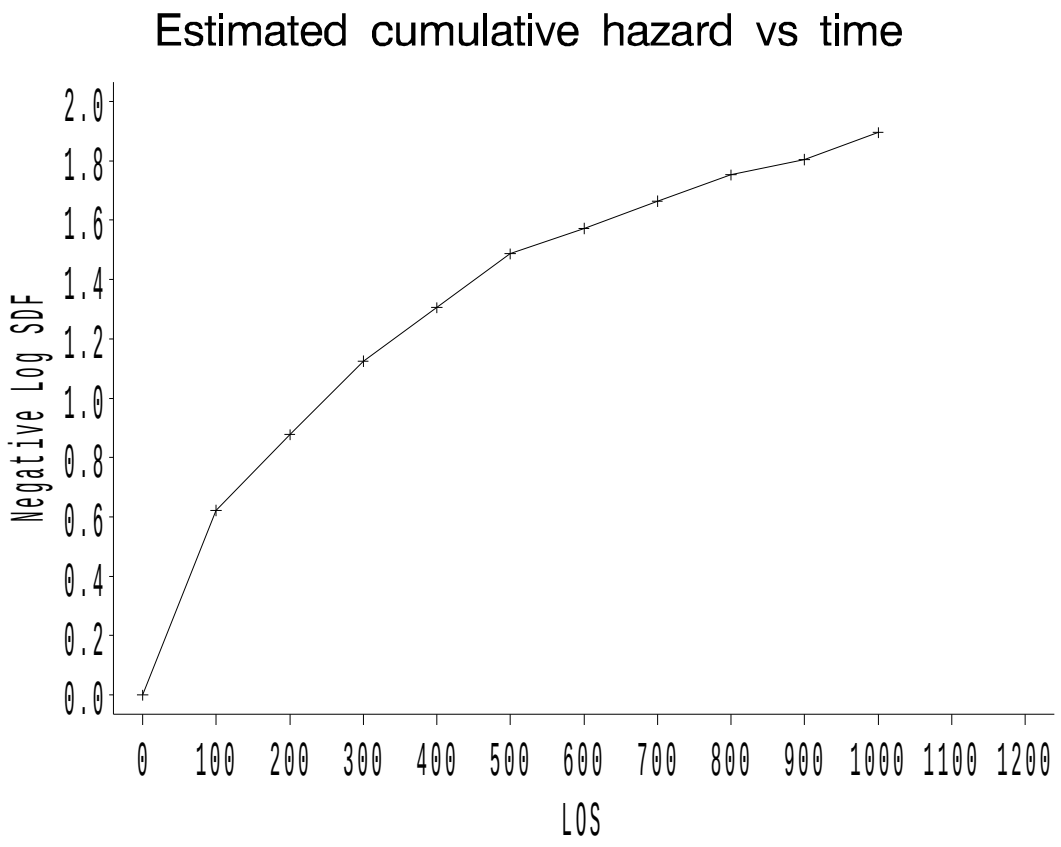
If T is Weibull, then $S(t) = \exp(-(\lambda t)^{\kappa})$

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

a straight line in $\log(t)$ with slope κ and intercept $\log(\lambda)$.

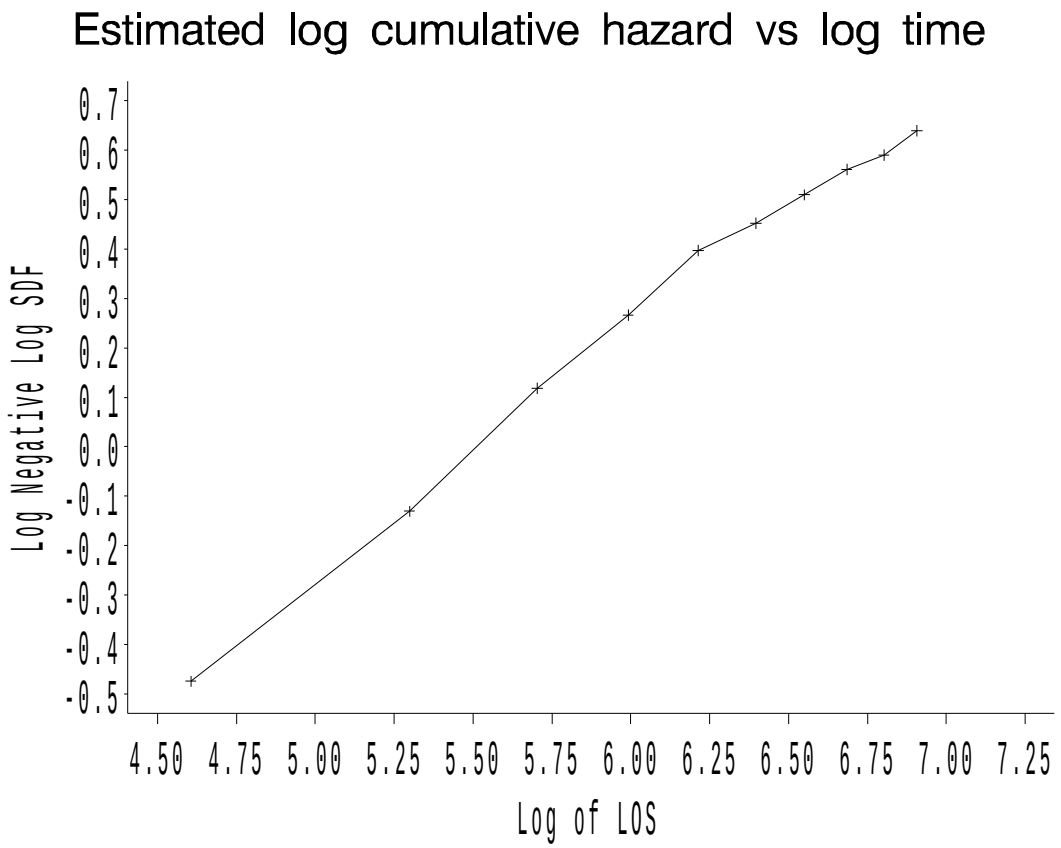
So we can calculate our estimated $\Lambda(t)$ and plot it versus t, and if it seems to form a straight line, then the exponential distribution is probably appropriate for our dataset.

Plots for nursing home data: $\hat{\Lambda}(t)$ vs t



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

Plots for nursing home data: $\log[-log(\hat{S}(t))]$ **vs** $\log(t)$



Comparison of Methods for the Two-sample problem:

Data:

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

In General:

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

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

MODELS

Exponential Regression:

$$\lambda_z(t) = \exp(\beta_0 + \beta_1 Z)$$

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

Weibull Regression:

$$\lambda_z(t) = \kappa \ \exp(\beta_0 + \beta_1 Z) \ t^{\kappa - 1}$$

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

Proportional Hazards Model:

$$\lambda_z(t) = \lambda_0(t) \exp(\beta_1)$$

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

Remarks

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

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

  How can you show this?

- If either the exponential model or the Weibull model is valid, then these models will tend to be more efficient than PH (smaller s.e.'s of estimates). This is because they assume a particular form for $\lambda_0(t)$, rather than estimating it at every death time.

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

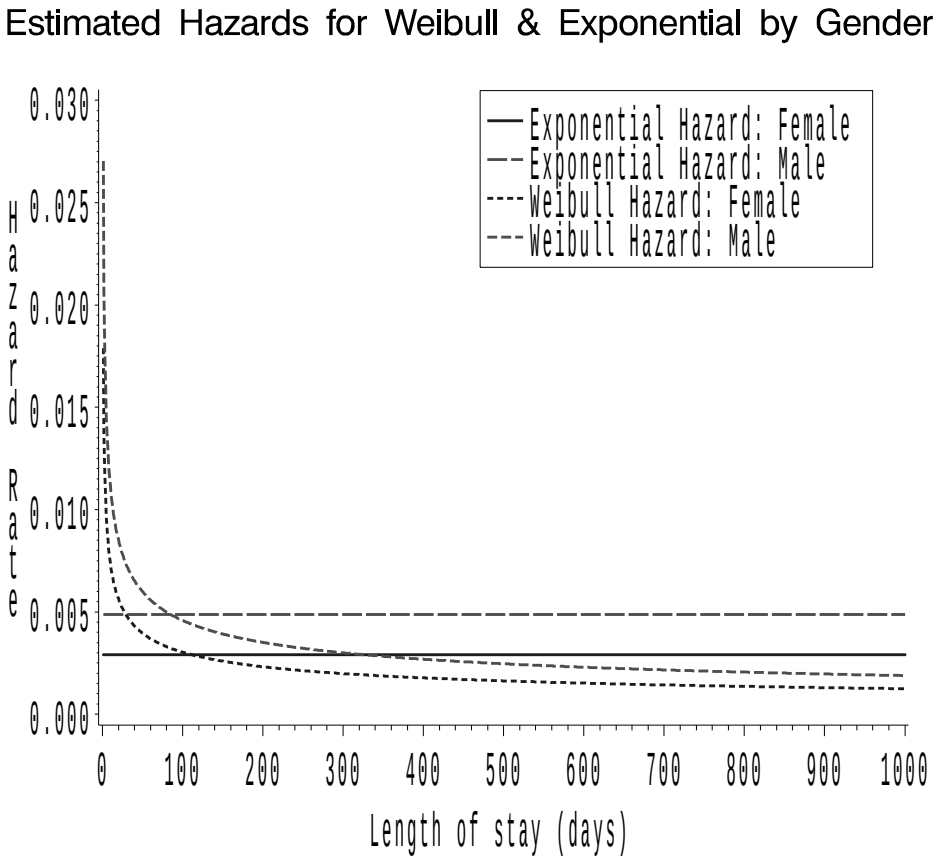
$$\begin{aligned} Z_i = 0 \Rightarrow \quad \hat{\lambda}_0 &= \exp(\hat{\beta}_0) \\ Z_i = 1 \Rightarrow \quad \hat{\lambda}_0 &= \exp(\hat{\beta}_0 + \hat{\beta}_1) \end{aligned}$$

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

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

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

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



Comparison with Proportional Hazards Model

```
. stcox gender, nohr

        failure _d:  fail
  analysis time _t:  los

Iteration 0:   log likelihood = -8556.5713
Iteration 1:   log likelihood = -8537.8013
Iteration 2:   log likelihood = -8537.5605
Iteration 3:   log likelihood = -8537.5604
Refining estimates:
Iteration 0:   log likelihood = -8537.5604

Cox regression -- Breslow method for ties

No. of subjects =         1591              Number of obs   =      1591
No. of failures =         1269
Time at risk    =       386211
                                            LR chi2(1)      =     38.02
Log likelihood  =   -8537.5604              Prob > chi2     =    0.0000

------------------------------------------------------------------------------
      _t |
      _d |      Coef.   Std. Err.     z     P>|z|  [95% Conf. Interval]
---------+--------------------------------------------------------------------
  gender |   .3943588   .0621004    6.350   0.000  .2726441    .5160734
------------------------------------------------------------------------------
```

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

Comparison with the Logrank and Wilcoxon Tests

```
. sts test gender

         failure _d:  fail
   analysis time _t:  los


Log-rank test for equality of survivor functions
------------------------------------------------

       |  Events
gender |  observed       expected
-------+-------------------------
0      |       902         995.40
1      |       367         273.60
-------+-------------------------
Total  |      1269        1269.00

             chi2(1) =      41.08
             Pr>chi2 =     0.0000


. sts test gender, wilcoxon

         failure _d:  fail
   analysis time _t:  los


Wilcoxon (Breslow) test for equality of survivor functions
----------------------------------------------------------

       |  Events                       Sum of
gender |  observed       expected       ranks
-------+---------------------------------------
0      |       902         995.40       -99257
1      |       367         273.60        99257
-------+---------------------------------------
Total  |      1269        1269.00            0

             chi2(1) =      41.47
             Pr>chi2 =     0.0000
```

Comparison of Hazard Ratios and Test Statistics for effect of Gender

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

Comparison of Mean and Median Survival Times by Gender

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

The Accelerated Failure Time Model

The general form of an accelerated failure time (AFT) model is:

$$\log(T_i) = \boldsymbol{\beta}_{AFT}\mathbf{Z}_i + \sigma\epsilon$$

where

- $\log(T_i)$ is the log of a survival time
- $\boldsymbol{\beta}_{AFT}$ is the vector of AFT model parameters corresponding to the covariate vector $\mathbf{Z}_i$
- ϵ is a random "error" term
- σ is a scale factor

In other words, we can model the log-survival times as a linear function of the covariates.

PROC LIFEREG in SAS and the `streg` command in STATA (without the exponential or weibull option) all use this "log-linear" model formulation for fitting parametric models.

By choosing different distributions for ϵ, we can obtain different parametric distributions:

- Exponential
- Weibull
- Gamma
- Log-logistic
- Normal
- Lognormal

We can compare the predicted survival under any of these parametric distributions to the KM estimated survival to see which one seems to fit best.

Once we decide on a certain class of model (say, Gamma), we can evaluate the contributions of covariates by finding the MLE's, and constructing Wald, Score, or LR tests of the covariate effects.

We can motivate the AFT model by first demonstrating the following two relationships:

- **1. For the Exponential Model:**

  If the failure times $T_i = T(\mathbf{Z}_i)$ follow an exponential distribution, i.e., $S_i(t) = e^{-\lambda_i t}$ with $\lambda_i = exp(\boldsymbol{\beta}\mathbf{Z}_i)$, then

  $$\log(T_i) \;=\; -\boldsymbol{\beta}\mathbf{Z}_i + \epsilon$$

  where ϵ follows an extreme value distribution (which just means that e^{ϵ} follows a unit exponential distribution).

- **2. For the Weibull Model:**

  If the failure times $T_i = T(\mathbf{Z}_i)$ follow a Weibull distribution, i.e., $S_i(t) = e^{\lambda_i t^{\kappa}}$ with $\lambda_i = exp(\boldsymbol{\beta}\mathbf{Z}_i)$, then

  $$\log(T_i) \;=\; -\sigma\boldsymbol{\beta}\mathbf{Z}_i + \sigma\epsilon$$

  where ϵ again follows an extreme value distribution, and $\sigma = 1/\kappa$.

In other words, both the Exponential and Weibull model can be written in the form of a log-linear model for the survival times, if we choose the right distribution for ϵ.

The log-linear form for the exponential can be derived by:

(1) Creating a new variable $T_0 = T_Z \times \exp(\boldsymbol{\beta}\mathbf{Z}_i)$

(2) Taking the log of T_Z, yielding $\log(T_Z) = \log\left(\frac{T_0}{\exp(\boldsymbol{\beta}\mathbf{Z}_i)}\right)$

Step (1): For an exponential model, recall that:

$$S_i(t) = Pr(T_Z \geq t) = e^{-\lambda t}, \quad \text{with } \lambda = \exp(\boldsymbol{\beta}\mathbf{Z}_i)$$

It follows that $T_0 \sim exp(1)$:

$$\begin{aligned} S_0(t) = Pr(T_0 \geq t) &= Pr(T_Z \cdot \exp(\boldsymbol{\beta}\mathbf{Z}) \geq t) \\ &= Pr(T_Z \geq t \exp(-\boldsymbol{\beta}\mathbf{Z})) \\ &= \exp\left[-\lambda\ t\ \exp(-\boldsymbol{\beta}\mathbf{Z})\right] \\ &= \exp\left[-\exp(\boldsymbol{\beta}\mathbf{Z})\ t\ \exp(-\boldsymbol{\beta}\mathbf{Z})\right] \\ &= \exp(-t) \end{aligned}$$

Step (2): Now take the log of the survival time:

$$\begin{aligned} \log(T_Z) &= \log\left(\frac{T_0}{\exp(\boldsymbol{\beta}\mathbf{Z}_i)}\right) \\ &= \log(T_0) - \log\left(\exp(\boldsymbol{\beta}\mathbf{Z}_i)\right) \\ &= -\boldsymbol{\beta}\mathbf{Z}_i + log(T_0) \\ &= -\boldsymbol{\beta}\mathbf{Z}_i + \epsilon \end{aligned}$$

where $\epsilon = \log(T_0)$ follows the **extreme value** distribution.

Relationship between Exponential and Weibull

If T_Z has a Weibull distribution, i.e., $S(t) = e^{-\lambda t^\kappa}$
with $\lambda = \exp(\boldsymbol{\beta}\mathbf{Z}_i)$, then you can show that the new variable

$$T_Z^* = T_Z^\kappa$$

follows an exponential distribution with parameter $\exp(\boldsymbol{\beta}\mathbf{Z}_i)$.
Based on the previous page, we can therefore write:

$$\log(T^*) = -\boldsymbol{\beta}\mathbf{Z} + \epsilon$$

(where ϵ has an extreme value distribution.)

But since $\log(T^*) = \log(T^\kappa) = \kappa \times \log(T)$, we can write:

$$\begin{aligned} \log(T) &= \log(T^*)/\kappa \\ &= (1/\kappa)\,(-\boldsymbol{\beta}\mathbf{Z}_i + \epsilon) \\ &= -\sigma\boldsymbol{\beta}\mathbf{Z}_i + \sigma\epsilon \end{aligned}$$

where $\sigma = 1/\kappa$.

This motivates the following general definition of the **Accelerated Failure Time Model** by:

$$\log(T_i) = \boldsymbol{\beta}_{AFT}\mathbf{Z}_i + \sigma\epsilon$$

where ϵ is a random "error" term, σ is a scale factor, Y is the log of a survival random variable, and

$$\boldsymbol{\beta}_{AFT} = -\sigma\boldsymbol{\beta}_e$$

where $\boldsymbol{\beta}_e$ came from the hazard $\lambda = \exp(\boldsymbol{\beta}\mathbf{Z})$.

The defining feature of an AFT model is:

$$S(t;\mathbf{Z}) = S_i(t) = S_0(\phi\ t)$$

That is, the effect of covariates is to accelerate (stretch) or decelerate (shrink) the time-scale.

Effect of AFT on hazard:

$$\lambda_i(t) \ = \ \phi\ \lambda_0(\phi t)$$

One way to interpret the AFT model is via its effect on median survival times. If $S_i(t) = 0.5$, then $S_0(\phi t) = 0.5$. This means:

$$M_i \;=\; \phi M_0$$

Interpretation:

- For $\phi < 1$, there is an acceleration of the endpoint (if $M_0 = 2$yrs in control and $\phi = 0.5$, then $M_i = 1$yr.
- For $\phi > 1$, there is a stretching or delay in endpoint
- In general, the lifetime of individual i is ϕ times what they would have experienced in the reference group

Since ϕ must be positive and a function of the covariates, we model $\phi = \exp(\boldsymbol{\beta}\mathbf{Z}_i)$.

When does Proportional hazards = AFT?

According to the proportional hazards model:

$$S(t) = S_0(t)^{\exp(\boldsymbol{\beta}\mathbf{Z}_i)}$$

and according to the accelerated failure time model:

$$S(t) = S_0(t \exp(\boldsymbol{\beta}\mathbf{Z}_i))$$

Say $T_i \sim Weibull(\lambda, \kappa)$. Then $\lambda(t) = \lambda\kappa t^{(\kappa-1)}$

Under the AFT model:

$$\begin{aligned}
\lambda_i(t) &= \phi\ \lambda_0(\phi t) \\
&= e^{\boldsymbol{\beta}\mathbf{Z}_i}\ \lambda_0(e^{\boldsymbol{\beta}\mathbf{Z}_i} t) \\
&= e^{\boldsymbol{\beta}\mathbf{Z}_i}\ \lambda_0\kappa \left(e^{\boldsymbol{\beta}\mathbf{Z}_i} t\right)^{(\kappa-1)} \\
&= \left(e^{\boldsymbol{\beta}\mathbf{Z}_i}\right)^{\kappa}\ \lambda_0\kappa t^{(\kappa-1)} \\
&= \left(e^{\boldsymbol{\beta}\mathbf{Z}_i}\right)^{\kappa}\ \lambda_0(t)
\end{aligned}$$

But this looks just like the PH model:

$$\lambda_i(t) = \exp(\boldsymbol{\beta}^*\mathbf{Z}_i)\ \lambda_0(t)$$

It turns out that the Weibull distribution (and exponential, since this is just a special case of a Weibull with $\kappa = 1$) is the only one for which the accelerated failure time and proportional hazards models coincide.

Special cases of AFT models

- Exponential regression: $\sigma = 1$, ϵ following the extreme value distribution.
- Weibull regression: σ arbitrary, ϵ following the extreme value distribution.
- Lognormal regression: σ arbitrary, ϵ following the normal distribution.

Examples in stata: Using the STREG command, one has the following options of distributions for the log-survival times:

```
. streg trt, dist(lognormal)
```

- exponential
- weibull
- gompertz
- lognormal
- loglogistic
- gamma

```
. streg gender, dist(exponential) nohr
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gender |    .516186   .0619148      8.337   0.000       .3948352    .6375368
------------------------------------------------------------------------------

. streg gender, dist(weibull) nohr
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gender |   .4138082   .0621021      6.663   0.000       .2920903    .5355261
         1/p |   1.627501   .0377726                          1.555127    1.703243
------------------------------------------------------------------------------

. streg gender, dist(lognormal)
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gender |  -.6743434   .1127352     -5.982   0.000      -.8953002   -.4533866
       _cons |   4.957636   .0588939     84.179   0.000       4.842206    5.073066
       sigma |    1.94718    .040584                           1.86924    2.028371
------------------------------------------------------------------------------

. streg gender, dist(gamma)
------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.       z     P>|z|       [95% Conf. Interval]
-------------+----------------------------------------------------------------
      gender |  -.6508469   .1147116     -5.674   0.000      -.8756774   -.4260163
       _cons |   4.788114   .1020906     46.901   0.000        4.58802    4.988208
       sigma |    1.97998   .0429379                          1.897586    2.065951
------------------------------------------------------------------------------
```

This gives a good idea of the sensitivity of the test of gender to the choice of model. It is also easy to get predicted survival curves under any of the parametric models using the following:

```
. streg gender, dist(gamma)
. stcurv, survival
```

The options HAZARD and CUMHAZ can also be substituted for SURVIVAL above to obtain plots.

AFT models in SAS

```
proc lifereg  data=pop covout  outest=survres;

  model los*censor(1)=gender /dist=exponential;

  model los*censor(1)=gender /dist=weibull;

  model los*censor(1)=gender /dist=gamma;

  model los*censor(1)=gender /dist=normal;
```

Other options are lognormal, logistic, and log-logistic. The default is to model log of response. Can specify "NOLOG" for no log-transformation. In this case, "normal" is the same as "lognormal."

Lifereg Procedure

Log Likelihood for EXPONENT -3320.476626

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	5.84213388	0.033296	30786	0.0001	Intercept
GENDER	1	-0.5161878	0.061915	69.50734	0.0001	
SCALE	0	1	0			Extreme value scale

Lagrange Multiplier ChiSquare for Scale 337.5998 Pr>Chi is 0.0001.

Log Likelihood for WEIBULL -3045.276811

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	5.75644118	0.0542	11280.04	0.0001	Intercept
GENDER	1	-0.6734732	0.101067	44.40415	0.0001	
SCALE	1	1.62750085	0.037773			Extreme value scale

Log Likelihood for GAMMA -2970.388508

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	4.78811071	0.104333	2106.114	0.0001	Intercept
GENDER	1	-0.6508468	0.114748	32.17096	0.0001	
SCALE	1	1.97998063	0.043107			Gamma scale parameter
SHAPE	1	-0.1906006	0.094752			Gamma shape parameter

Log Likelihood for NORMAL -9593.512838

Variable	DF	Estimate	Std Err	ChiSquare	Pr>Chi	Label/Value
INTERCPT	1	303.824624	9.919629	938.1129	0.0001	Intercept
GENDER	1	-107.09585	18.97784	31.84577	0.0001	
SCALE	1	330.093584	6.918237			Normal scale parameter

Designing a Survival Study

We will focus on the power of tests based on the exponential distribution and the logrank test.

- As in standard designs, the power depends on
  - The Type I error (significance level)
  - The difference of interest, Δ, under H_a.
- A notable difference from the usual scenario is that power depends on the **number of failures**, not the total sample size.
- In practice, designing a survival study involves deciding how many patients or individuals to enter, as well as how long they should be followed.
- Designs may be **fixed sample size** or **sequential** (More on this later!)

References:

Collett	Chapter 12
Pocock	Chapter 9 of *Clinical Trials*
Williams	Chapter 10 of *AIDS Clinical Trials* (eds. Finkelstein and Schoenfeld)

Review of power calculations for 2-sample normal

Suppose we have the following data:

$$\begin{aligned} &\text{Group 1:} \qquad (Y_{11}, \ldots Y_{1n_1}) \\ &\text{Group 0:} \qquad (Y_{01}, \ldots Y_{0n_0}) \end{aligned}$$

and make the following assumptions:

$$Y_{1j} \sim \mathcal{N}(\mu_1, \sigma^2) \qquad Y_{0j} \sim \mathcal{N}(\mu_0, \sigma^2)$$

Our objective is to test:

$$H_0 : \mu_1 = \mu_0 \quad \Rightarrow \quad H_0 : \triangle = 0 \qquad \text{where } \triangle = \mu_1 - \mu_0$$

The standard test is based on the Z statistic:

$$Z = \frac{\overline{Y_1} - \overline{Y_0}}{\sqrt{s^2(\frac{1}{n_1} + \frac{1}{n_0})}}$$

where s^2 is the pooled sample variance (we are assuming equal variances here). This test statistic follows a $\mathcal{N}(0, 1)$ distribution under H_0.

If the sample sizes are equal in the two arms, $n_0 = n_1 = n/2$, (which will *maximize* the power), then we have the simpler form:

$$Z \;=\; \frac{\overline{Y_1} - \overline{Y_0}}{\sqrt{s^2(\frac{1}{n/2} + \frac{1}{n/2})}} = \frac{\overline{Y_1} - \overline{Y_0}}{2s/\sqrt{n}}$$

The steps to follow in calculating the sample size are:

(1) Determine the critical value, c, for rejecting the null when it is true.

(2) Calculate the probability of rejecting the null when the alternative is true, substituting c from above.

(3) Rewrite the expression in terms of the sample size for a given power.

Step (1):

Set the significance level, α, equal to the probability of rejecting the null hypothesis when it is true:

$$\begin{aligned}
\alpha &= Pr\left(|\overline{Y_1} - \overline{Y_0}| > c \mid H_0\right) \\
&= Pr\left(\frac{|\overline{Y_1} - \overline{Y_0}|}{2s/\sqrt{n}} > \frac{c}{2s/\sqrt{n}} \mid H_0\right) \\
&= Pr\left(|Z| > \frac{c}{2s/\sqrt{n}}\right) \;=\; 2 \cdot \Phi\left(\frac{c}{2s/\sqrt{n}}\right)
\end{aligned}$$

$$\text{so} \quad z_{1-\alpha/2} = \frac{c}{2s/\sqrt{n}}$$

$$\text{or} \quad c = \frac{z_{1-\alpha/2}\; 2\; s}{\sqrt{n}}$$

Note that z_γ is the value such that $\Phi(z_\gamma) = Pr(Z < z_\gamma) = \gamma$.

Step (2):

Calculate the probability of rejecting the null when H_a is true. Start out by writing down the probability of a Type II error:

$$\begin{aligned} \beta &= Pr\left(\text{accept } H_0 \mid H_a\right) \\ \text{so} \quad 1-\beta &= Pr\left(\text{reject } H_0 \mid H_a\right) \\ &= Pr\left(|\overline{Y_1} - \overline{Y_0}| > c \mid H_a\right) \\ &= Pr\left(\frac{|\overline{Y_1} - \overline{Y_0}| - \Delta}{2s/\sqrt{n}} > \frac{c-\Delta}{2s/\sqrt{n}} \mid H_a\right) \\ &= Pr\left(Z > \frac{c-\Delta}{2s/\sqrt{n}}\right) \end{aligned}$$

$$\text{so we get} \quad z_\beta = -z_{1-\beta} = \frac{c-\Delta}{2s/\sqrt{n}}$$

Now we substitute c from Step (1):

$$\begin{aligned} -z_{1-\beta} &= \frac{z_{1-\alpha/2}\ 2\ s/\sqrt{n} - \Delta}{2s/\sqrt{n}} \\ &= z_{1-\alpha/2} - \frac{\Delta}{2\ s\ /\sqrt{n}} \end{aligned}$$

Step (3):

Now rewrite the equation in terms of sample size for a given power, $1-\beta$, and significance level, α:

$$\begin{aligned} z_{1-\alpha/2} + z_{1-\beta} &= \frac{\Delta}{2s/\sqrt{n}} \\ &= \frac{\Delta\sqrt{n}}{2s} \\ \Longrightarrow \quad n &= \frac{(z_{1-\alpha/2} + z_{1-\beta})^2 4s^2}{\Delta^2} \end{aligned}$$

Notes:

The power is an increasing function of the standardized difference:

$$\mu_T(\triangle) = \frac{\triangle}{2s/\sqrt{n}}$$

This is just the number of standard errors between the two means, under the assumption of equal variances.

1. As n increases, the power increases.
2. For fixed n, the power increases with $\triangle$.
3. For fixed n and $\triangle$, the power decreases with s.
4. Assigning equal numbers of patients to the two groups ($n_1 = n_0 = n/2$) is best in terms of maximizing power.

An Example:

$$n = \frac{\left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2 4s^2}{\triangle^2}$$

Say we want to derive the total sample size required to yield 90% power for detecting a difference of 0.5 standard deviations between means, based on a two-sided 0.05 level test.

$$\begin{aligned} \alpha &= 0.05 \\ z_{1-\frac{\alpha}{2}} &= 1.96 \\ \beta &= 0.10 \\ z_{1-\beta} = z_{0.90} &= 1.28 \end{aligned}$$

$$n = \frac{(1.96 + 1.28)^2 4s^2}{\triangle^2} \approx \frac{42\ s^2}{\triangle^2}$$

For a 0.5 standard deviation difference, $\Delta/s = 0.5$, so

$$n \approx \frac{42}{(0.5)^2} = 168$$

If you end up with $n < 30$, then you should be using the t-distribution rather than the normal to calculate critical values, and then the process is iterative.

Survival Studies: Comparing Proportions of Events

In some cases, the sample size for a survival trial is based on a crude comparison of the proportion of events at some fixed point in time.

In this case, we can apply the results just shown to get sample sizes, based on the normal approximation to the binomial:

Define:

P_c probability of event in control arm by time t

P_e probability of event in "experimental" arm by time t

The number of patients required per treatment arm based on a chi-square test comparing binomial proportions is:

$$N = \frac{\{z_{1-\frac{\alpha}{2}}\sqrt{2\overline{P}(1-\overline{P})} + z_{1-\beta}\sqrt{P_e(1-P_e)+P_c(1-P_c)}\}^2}{(P_c-P_e)^2}$$

where $\overline{P} = (P_e + P_c)/2$

(This looks slightly different because the variance is not the same under H_o and H_a, as was the case in the normal previous example.)

Notes on comparing proportions of failures:

- Use of chi-square test is best when $0.2 < P_e, P_c < 0.8$
- Should have ≥ 15 patients in each cell of the (2x2) table
- For smaller sample sizes, use Fisher's exact test to motivate power calculations
- Efficiency vs logrank test is near 100% for studies with short durations relative to the median event time

  What does this mean in terms of the event rates? High or low?

- Calculation of sample size for comparing proportions often provides an upper bound to those based on comparison of survival distributions

Sample size based on the logrank test

Recap: Consider a two group survival problem, with equal numbers of individuals in the two groups (say n_0 in group 0 and n_1 in group 1). Let $\tau_1, ..., \tau_K$ represent the K ordered, distinct failure times, and at the j-th event time:

	Die/Fail		
Group	Yes	No	Total
0	d_{0j}	$r_{0j} - d_{0j}$	r_{0j}
1	d_{1j}	$r_{1j} - d_{1j}$	r_{1j}
Total	d_j	$r_j - d_j$	r_j

where d_{0j} and d_{1j} are the number of deaths (events) in group 0 and 1, respectively, at the j-th event time, and r_{0j} and r_{1j} are the corresponding numbers at risk.

The logrank test is: (z-statistic version)

$$Z_{LR} = \frac{\Sigma_{j=1}^{K}(d_{1j} - e_j)}{\sqrt{\Sigma_{j=1}^{K} v_j}}$$

$$\text{with} \quad e_j = d_j \, r_{1j}/r_j$$

$$v_j = r_{1j} r_{0j} d_j (r_j - d_j)/[r_j^2 (r_j - 1)]$$

Distribution of the logrank statistic

Suppose that the hazard rates in the two groups are $\lambda_0(t)$ and $\lambda_1(t)$, with hazard ratio

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

and suppose we are interested in testing $H_o : \beta = ln(\theta) = 0$ (which is equivalent to testing $H_o : \theta = 1$.)

[Note: we will use $ln(\theta)$ rather than β in the following, so that there is no confusion with the Type II error rate]

It is possible to show that

- if there are no ties, and
- we are "near" H_0 :

then:

- $E(d_{1j} - e_j | d_{1j}, d_{0j}, r_{1j}, r_{0j}) \approx ln(\theta)/4$
- $v_j \approx 1/4$

So, at a point $ln(\theta)$ in the alternative, we get:

$$Z_{LR} \approx \frac{\Sigma_{j=1}^{K} ln(\theta)/4}{\sqrt{\Sigma_{j=1}^{K} 1/4}} = \frac{d\ ln(\theta)/4}{\sqrt{d/4}} = \frac{\sqrt{d}\ ln(\theta)}{2}$$

and $Z_{LR} \sim N(\ ln(\theta)\sqrt{d}/2\ ,\ 1\ )$

Heuristic Proof:

$$\begin{aligned} E(d_{1j}|d_{1j}, d_{0j}, r_{1j}, r_{0j}) &= Pr(d_{1j} = 1|d_j = 1, r_{1j}, r_{0j}) \\ &= \frac{r_{1j}\lambda_0\theta}{r_{1j}\lambda_0\theta + r_{0j}\lambda_0} \\ &= \frac{r_{1j}\theta}{r_{1j}\theta + r_{0j}} \\ &= \frac{r_{1j}}{r_{1j} + r_{0j}} + ln(\theta)\left[\frac{r_{1j}r_{0j}}{(r_{1j} + r_{0j})^2}\right] \end{aligned}$$

But $e_j = r_{1j}/(r_{1j} + r_{0j})$, so:

$$E(d_{1j}|d_{1j}, d_{0j}, r_{1j}, r_{0j}) - e_j = ln(\theta)\left[\frac{r_{1j}r_{0j}}{(r_{1j} + r_{0j})^2}\right]$$

If $n_0 = n_1$, then near H_0 :, $r_{1j} \approx r_{0j}$, hence,

$$E(d_{1j}|d_{1j}, d_{0j}, r_{1j}, r_{0j}) - e_j = ln(\theta)/4$$

Similarly, with no ties, we have

$$v_j = r_{1j}r_{0j}/r_j^2 \approx 1/4$$

This can also be derived via the partial likelihood:

We can write the partial likelihood as:

$$\begin{aligned} l(\boldsymbol{\beta}) &= \log\left[\prod_{j=1}^{n}\left(\frac{e^{\boldsymbol{\beta}\mathbf{Z}_j}}{\sum_{\ell\in\mathcal{R}(\tau_j)} e^{\boldsymbol{\beta}\mathbf{Z}_\ell}}\right)^{\delta_j}\right] \\ &= \sum_{j=1}^{n}\delta_j\left[\boldsymbol{\beta}\mathbf{Z}_j - \log\left(\sum_{\ell\in\mathcal{R}(\tau_j)} e^{\boldsymbol{\beta}\mathbf{Z}_\ell}\right)\right] \end{aligned}$$

and then the **"score"** (partial derivative of log-likelihood) becomes:

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

We can write the **"information"** (minus second partial derivative of the log-likelihood) as:

$$-\frac{\partial^2}{\partial\boldsymbol{\beta}^2}\ell(\boldsymbol{\beta}) = \sum_{j=1}^{n}\delta_j\left[\frac{\sum_{\ell\in\mathcal{R}(\tau_j)} e^{\boldsymbol{\beta}\mathbf{Z}_\ell}\sum_{\ell\in\mathcal{R}(\tau_j)}\mathbf{Z}_\ell e^{\boldsymbol{\beta}\mathbf{Z}_\ell} - (\sum_{\ell\in\mathcal{R}(\tau_j)}\mathbf{Z}_\ell e^{\boldsymbol{\beta}\mathbf{Z}_\ell})^2}{\sum_{\ell\in\mathcal{R}(\tau_j)} e^{\boldsymbol{\beta}\mathbf{Z}_\ell}}\right]$$

The logrank statistic (with no ties) is equivalent to the score statistic for testing $\boldsymbol{\beta} = 0$:

$$Z_{LR}\ \ \frac{U(0)}{\sqrt{I(0)}}$$

By a Taylor series expansion:

$$U(0) \cong U(\beta) - \beta\frac{\partial U}{\partial\beta}(0)$$

$$E[U(0)] \cong \beta d/4 \quad \text{and} \quad I(0) \cong d/4$$

Power of the Logrank Test

Using a similar argument to before, the power of the logrank test (based on a two-sided α level test) is approximately:

$$\text{Power}(\theta) \approx 1 - \Phi\left[z_{1-\frac{\alpha}{2}} - ln(\theta)\sqrt{d}/2\right]$$

Note: Power depends only on d and θ!

We can easily solve for the required number of events to achieve a certain power at a specified value of θ:

To yield power$(\theta) = 1 - \beta$, we want d so that

$$\begin{aligned} 1-\beta &= 1 - \Phi\left(z_{1-\frac{\alpha}{2}} - ln(\theta)\sqrt{d}/2\right) \\ \Rightarrow \quad z_\beta &= z_{1-\frac{\alpha}{2}} - ln(\theta)\sqrt{d}/2 \\ \Rightarrow \quad d &= \frac{4\left(z_{1-\frac{\alpha}{2}} - z_\beta\right)^2}{[ln(\theta)]^2} \\ \text{or} \quad d &= \frac{4\left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2}{[ln(\theta)]^2} \end{aligned}$$

Example:

Say we were planning a 2-arm study, and wanted to be able to detect a hazard ratio of 1.5 with 90% power at a 2-sided significance level of $\alpha = 0.05$.

Required number of events:

$$\begin{aligned} d &= \frac{4\left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2}{[ln(\theta)]^2} \\ &= \frac{4(1.96 + 1.282)^2}{[ln(1.5)]^2} \\ &\approx \frac{42}{0.1644} = 256 \end{aligned}$$

Events required for various Hazard Ratios

Hazard	**Power**	
Ratio	80%	90%
1.5	191	256
2.0	66	88
2.5	38	50
3.0	26	35

Most studies are designed to detect a hazard ratio of 1.5-2.0.

Practical Considerations

- How do we decide on θ?
- How do we translate numbers of failures to numbers of patients?

Hazard ratios for the exponential distribution

The hazard ratio from two exponential distributions can be easily translated into more intuitively interpretable quantities:

Median:

If $T_i \sim exp(\lambda_i)$, then

$$\text{Median}(T_i) = -ln(0.5)/\lambda_i$$

It follows that

$$\frac{\text{Median}(T_1)}{\text{Median}(T_0)} = \frac{\lambda_0}{\lambda_1} = e^{-\beta} = \frac{1}{\theta}$$

Hence, doubling the median survival of a treated compared to a control group will correspond to halving the hazard.

R-year survival rates

Suppose the R-year survival rate in group 1 is $S_1(R)$ and in group 0 is $S_0(R)$. Under the exponential model:

$$S_i(R) = exp(-\lambda_i R)$$

Hence,

$$\frac{ln(S_1(R))}{ln(S_0(R))} = \frac{-\lambda_1 R}{-\lambda_0 R} \;=\; \frac{\lambda_1}{\lambda_0} \;=\; e^{\beta} \;=\; \theta$$

Hence, doubling the hazard rate from group 1 to group 0 will correspond to doubling the log of the R-year survival rate. Note that this result does not depend on R!.

Example: Suppose the 5-year survival rate on treatment A is 20% and we want 90% power to detect an improvement of that rate to 30%. The corresponding hazard ratio of treated to control is:

$$\frac{ln(0.3)}{ln(0.2)} = \frac{-1.204}{-1.609} = 0.748$$

From our previous formula, the number of events (deaths) needed to detect this improvement with 90% power, based on a 2-sided 5% level test is:

$$d = \frac{4(1.96 + 1.282)^2}{[ln(0.748)]^2} = 499$$

Translating to Number of Enrolled Patients

First, suppose that we will enter N patients into our study at time 0, and will then continue the study for F units of time.

Under H_0, the probability that an individual will fail during the study is:

$$\begin{aligned} Pr(fail) &= \int_0^F \lambda_0 e^{-\lambda_0 t} dt \\ &= 1 - e^{-\lambda_0 F} \end{aligned}$$

Hence, if our calculations say we need d failures, then to decide how many patients to enter, we simply solve

$$d = (N/2)(1 - e^{-\lambda_0 F}) + (N/2)(1 - e^{-\lambda_1 F})$$

To solve the above equation for N, we need to supply values of F and d. In other words, here we are already deciding what HR we want to detect (with what power, etc), and for how long we are going to follow patients. What we get is the total number of patients we need to enroll in order to observe the desired number of events in F units of follow-up time.

Example: Suppose we want to detect a 50% improvement in the median survival from 12 months to 18 months with 80% power at $\alpha = 0.05$, and we plan on following patients for 3 years (36 months).

We can use the two medians to calculate both the parameters λ_0 and λ_1 and the hazard ratio, θ:

$$\begin{aligned}
\text{Median}(T_i) &= -ln(0.5)/\lambda_i \\
\text{so } \lambda_1 &= \frac{-ln(0.5)}{M1} = \frac{0.6931}{18} = 0.0385 \\
\lambda_0 &= \frac{-ln(0.5)}{M0} = \frac{0.6931}{12} = 0.0578 \\
\theta &= \frac{\lambda_1}{\lambda_0} = \frac{0.0385}{0.0578} = \frac{12}{18} = 0.667
\end{aligned}$$

and from our previous table, # events required is $d = 191$ (same for $\theta = 1.5$ as it is for 1/1.5=0.667).

So we need to solve:

$$\begin{aligned}
191 &= (N/2)(1 - e^{-0.0578*36}) + (N/2)(1 - e^{-0.0385*36}) \\
&= (N/2)(0.875) + (N/2)(0.7500) = (N/2)(1.625) \\
\Rightarrow \quad N &= 235
\end{aligned}$$

(for practical reasons, we would probably round up to 236 and randomize 118 patients to each treatment arm)

A more realistic accrual pattern

In reality, not everyone will enter the study on the same day. Instead, the accrual will occur in a "staggered" manner over a period of time.

The standard assumption:
Suppose individuals enter the study uniformly over an accrual period lasting A units of time, and that after the accrual period, follow-up will continue for another F units of time.

To translate d to N, we need to calculate the probability that a patient fails under this accrual and follow-up scenario.

$$\begin{aligned} Pr(\text{fail}) &= \int_0^A Pr(\text{fail}|\text{enter at } a)\ f(a)\ da \\ &= 1 - \frac{\int_0^A S(a+F)\ da}{A} \end{aligned} \tag{2}$$

$$\begin{aligned} \text{Then solve:} \quad d &= (N/2)Pr(\text{fail};\lambda_0) + (N/2)Pr(\text{fail};\lambda_1) \\ &= (N/2)P_c + (N/2)P_e \\ &= (N/2)(P_c + P_e) \end{aligned}$$

If we now solve for N (substituting in formula for d), we get:

$$N = \frac{2\ d}{(P_c + P_e)}$$

$$N = \frac{8\ \left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2}{[ln(\theta)]^2}\ \ \frac{1}{(P_c + P_e)}$$

How can we get P_c and P_e from (2)?

If we assume that the exponential distribution holds, then we can solve (2) to obtain:

$$P_i \;=\; 1 - \frac{exp(-\lambda_i F)(1 - exp(-\lambda_i A))}{\lambda_i A} \qquad (3)$$

(for $i = c, e$)

Freedman suggested an approximation for P_c and P_e, by computing the probability of an event at the median duration of follow-up, $(A/2 + F)$:

$$P_i = Pr(\text{fail}; \lambda_i) \;=\; 1 - exp[-\lambda_i (A/2 + F)] \qquad (4)$$

He showed that this approximation works pretty well for the exponential distribution (i.e., it gives values close to (3)).

An alternative formulation

Rubenstein, Gail, and Santner (1981) suggest the following approach for calculating the total sample size that must be enrolled:

$$N \;=\; \frac{2\;\left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2}{[ln(\theta)]^2}\;\left[\frac{1}{P_c} + \frac{1}{P_e}\right]$$

where P_c and P_e are the expected proportion of patients or individuals who will fail (have an event) on the control and treatment arms.

How do we calculate (estimate) P_c and P_e ?

- using the general formula for a distribution S given in (2)
- using the exact formula for an exponential distribution given in (3)
- using the approximation given by (4)

Note: all of these formulas can be modified for **unequal** assignment to treatment (or exposure) groups by changing $(N/2)$ in the formulas on p.17-19 to $(qc * N)$ and $(qe * N)$, where qc and qe are the proportions assigned to the control and exposed groups, respectively.

Freedman's Approach (1982)

Freedman's approach is based on the logrank statistic under the assumption of proportional hazards, but does not require the assumption of exponential survival distributions.

Total number of events:

$$d \;=\; \left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2 \quad \left(\frac{\theta+1}{\theta-1}\right)^2$$

Total sample size:

$$N \;=\; \frac{2\;\left(z_{1-\frac{\alpha}{2}} + z_{1-\beta}\right)^2}{P_e + P_c} \quad \left(\frac{\theta+1}{\theta-1}\right)^2$$

where P_e and P_c are estimated using (4).

This approximation depends on the assumption of a constant ratio between the number of patients at risk in the two treatment groups prior to each event time $\Longrightarrow r_{0j} \approx r_{1j}$ (as shown in the "heuristic proof"). When this assumption is not satisfied, the required sample sizes tend to be overestimated.

Q. When would this assumption not be satisfied?

A. When the smallest detectable difference is large.

Some examples of study design

Example I:

A clinical trial in esophageal cancer will randomize patients to radiotherapy alone (Rx A) versus radiotherapy plus chemotherapy (Rx B). The goal of the study is to compare the two treatments with respect to survival, and we plan to use the logrank test. From historical data, we know that the median survival on RX A for this disease is around 9 months. We want 90% power to detect an improvement in this median to 18 months. Past studies have been able to accrue approximately 50 patients per year. Choose a suitable study design.

Example II:

A clinical trial in early stage breast cancer will randomize patients after their surgery to Tamoxifen (ARM A) versus observation only (ARM B). The goal of the study is to compare the two treatments with respect to time to relapse, and the logrank test will be used in the analysis. From historical data, we know that after five years, 65% of the patients will still be disease free. We would like to have 90% power to detect an improvement in this disease free rate to 75%. Past studies have been able to accrue approximately 200 patients per year. Choose a suitable study design.

Example III:

Some investigators in the environmental health department want to conduct a study to assess the effects of exposure to toluene on time to pregnancy. They will conduct a cohort study involving women who work in a chemical factory in China. It is estimated that 20% of the women will have workplace exposure to toluene. Furthermore, it is known that among unexposed women, 80% will become pregnant within a year. The investigators will be able to enroll 200 women per year into the study, and plan an additional year of follow-up at the end of accrual. Assuming they have 2 years accrual, what reduction in the 1-year pregnancy rate for exposed women will they be able to detect with 85% power? What if they have 3 years of accrual?

Other important issues:

The approaches just described address the basic question of calculating a sample size for study with a survival endpoint.

These approaches often need to be modified slightly to address the following complications:

- **Loss to follow-up**
- **Non-compliance (or cross-overs)**
- **Stratification**
- **Sequential monitoring**
- **Equivalence hypotheses**

Next we summarize some of the main points.

Loss to follow-up

If some patients are lost to follow up (as opposed to censored at the end of the trial without the event), the power will be decreased.

There are two main approaches for dealing with this:

- **Simple inflation method** - If ℓ*100% of patients are anticipated to be lost to follow up, calculate target sample size to be

$$N^* = \left(\frac{1}{1-\ell}\right) \cdot N$$

  Example: Say you calculate $N = 200$, and anticipate losses of 20%. The simple inflation method would give you a target sample size of $N^* = (1/0.8) * 200 = 250$.

  Warning: people often make the mistake of just inflating the original sample size by ℓ*100%, which would have given $N* = 240$ for the example above.

- **Exponential loss assumption** - the above approach assumes that losses contribute **NO** information. But we actually have information on them up until the time that they are lost. Incorporate this by assuming that time to loss also follows an exponential distribution, and modify P_e and P_c.

Noncompliance

If some patients don't take their assigned treatments, the power will be decreased. This issue has two sides:

- **Drop-outs (d_e)** - patients who cannot tolerate the medication stop taking it; their hazard rate would become the same as the placebo group (if included in study) at that point.
- **Drop-ins (d_c)** - patients assigned to less effective therapy may not get relief from symptoms and seek other therapy, or request to cross-over.

A conservative remedy – adjust P_e and P_c as follows:

$$\begin{aligned} P_e^* &= P_e(1-d_e) + P_c\, d_e \\ P_c^* &= P_c(1-d_c) + P_e\, d_c \end{aligned}$$

Design Strategy:

1. Decide on
   - Type I error (significance level)
   - clinically important difference (in terms of HR)
   - desired power

2. Determine the number of failures needed

3. Based on past experience
   - decide on a reasonable distribution for the controls (usually exponential)
   - estimate anticipated accrual per unit time
   - estimate expected rate of loss to follow up

   Vary the values of A and F until you get something practically feasible that gives the right number of failures.

4. Consider noncompliance, sequential monitoring, and other issues impacting sample size

Included on the next several pages is a SAS program to calculate sample sizes for survival studies. It uses several of the approaches we've discussed, including:

- Rubenstein, Gail and Santner (RGS, 1981)
- Freedman (1982)
- Lachin and Foulkes (1986)

A copy of this program is shown on the next several pages. The program requires entry of:

- Significance level (alpha)
- Power
- Sides (1 for one-sided test, 2 for two-sided test)
- Accrual period
- Follow up period
- Yearly rate of loss to follow-up
- Proportion randomized to experimental treatment arm
- One of the following:
  - Yearly event rate on control and experimental treatment arms
  - Yearly event rate on control arm, and the hazard ratio
  - Median time to event on control and experimental treatment arms

The SAS program `rgsnew.sas`

```
data rgs;
  ************************************************************************;
  *** enter the following information in this block;
  alpha = 0.05;          /* significance level */
  sides = 2;             /* one-sided or two-sided test */
  power = 0.90;          /* Desired power */
  accrual = 2;           /* Accrual period in years */
  fu = 1.5;              /* Follow up after last patient is accrued */
  loss = 0.0;            /* yearly rate of loss */
  qe = 0.5;              /* proportion randomized to experimental arm */

  *** either enter the median time to event in years on control;
  *** or the yearly event rate - leave the other value missing;
  medianc = 0.75;        /* median time to event on control arm */
  probc = .;             /* yearly event rate in control arm */

  *** either enter the yearly event rate in the experimental arm  ;
  *** or the hazard ratio for control vs experimental             ;
  *** or the median time to event on experimental arm in years    ;
  *** leave the other values missing (.);
  mediane = 1.5;         /* median time to event on experimental */
  probe = .;             /* yearly event rate in experimental arm */
  rr=.;                  /* hazard ratio */
  ************************************************************************;
  beta = 1 - power;
  qc = 1 - qe;
  zalpha = probit(1-alpha/sides);
  zbeta = probit(1-beta);

  *** calculate yearly event rate in both arms using medians, if supplied;
  if medianc^=. then do;
    hazc=-log(0.5)/medianc;
    probc=1-exp(-hazc);
  end;
  if mediane^=. then do;
    haze=-log(0.5)/mediane;
    probe=1-exp(-haze);
  end;

  hazc = -log(1-probc);
```

```
*** calculate hazard in experimental group, using yearly event rate;
*** or hazard ratio;
if probe^=. then haze = -log(1-probe);
if rr^=. then haze = hazc/rr;

if probe^=. and haze^=. then do;
   put "************************************************************";
   put "WARNING: both yearly event rate and hazard ratio (HR) have";
   put "         been specified.  Calculations will use the HR";
   put "************************************************************" /;
end;

*** calculate median survival times if not supplied;
medianc=-log(0.5)/hazc;
mediane=-log(0.5)/haze;

hazl = -log(1-loss);
avghaz = qc*hazc + qe*haze;
rr = hazc/haze;
log_rr = log(hazc/haze);
totloss=(accrual*0.5 + fu)*loss;

*** compute expected probability of death (event) during trial;
*** given staggered accrual but NO loss;
pc0loss = 1-((exp(-hazc*fu)-exp(-hazc*(accrual+fu)))/(hazc*accrual));
pe0loss = 1-((exp(-haze*fu)-exp(-haze*(accrual+fu)))/(haze*accrual));

*** compute expected probability of event during trial;
*** given staggered accrual AND loss;
pc = (1 - (exp(-(hazc+hazl)*fu)-exp(-(hazc+hazl)*(fu+accrual)))
          /((hazc+hazl)*accrual))*(hazc/(hazc+hazl));
pe = (1 - (exp(-(haze+hazl)*fu)-exp(-(haze+hazl)*(fu+accrual)))
          /((haze+hazl)*accrual))*(haze/(haze+hazl));
pbar = (1 - (exp(-(avghaz+hazl)*fu)-exp(-(avghaz+hazl)*(fu+accrual)))
          /((avghaz+hazl)*accrual))*(avghaz/(avghaz+hazl));

*** compute total sample size assuming loss;
N = int(((zalpha+zbeta)**2)/(log_rr**2)*(1/(qc*pc)+1/(qe*pe))) + 1;

*** Compute sample size using method of Freedman (1982);
N_FRD = int((2*(((rr+1)/(rr-1))**2)*(zalpha+zbeta)**2)/(2*(qe*pe+qc*pc)))+1;
```

```
*** Compute sample size using method of Lachin and Foulkes (1986);
*** with rates under H0 given by pooled hazard;
N_LF = int((zalpha*sqrt((avghaz**2)*(1/pbar)*(1/qc + 1/qe)) +
       zbeta*sqrt((hazc**2)*(1/(qc*pc)) + (haze**2)*(1/(qe*pe))))**2/
       ((hazc-haze)**2)) + 1;

*** compute total sample size assuming no loss;
n_0loss = int(((zalpha+zbeta)**2)/(log_rr**2)*
           (1/(qc*pc0loss)+1/(qe*pe0loss))) + 1;

*** compute sample size using simple inflation method for loss;
naive = int(n_0loss/(1-totloss)) + 1;

*** verify that actual power is same as desired power;
newpower = probnorm(sqrt(((N)*(log_rr**2))/(1/(qc*pc) + 1/(qe*pe)))
            - zalpha);

if abs(newpower-power)>0.001 then do;
  put '*** WARNING: actual power is not equal to desired power';
  put 'Desired power: ' power ' Actual power: ' newpower;
end;

*********************************************************************;
*** compute number of events expected during trial;
*********************************************************************;
*** Compute expected number under null;
n_evth0 = int(n*pbar) + 1;

r=qc/qe;
*** Rubinstein, Gail and Santner (1981) method - simple approximation;
n_evtrgs = int((((r+1)**2)/r)*((zalpha + zbeta)**2)/(log_rr**2)) + 1;

*** Freedman (1982);
n_evtfrd = int((((rr+1)/(rr-1))**2) * (zalpha+zbeta)**2)+1;

*** Using backtracking method of Lachin and Foulkes (1986);
n_evt_c = int(N*qc*pc) + 1;
n_evt_e = int(N*qe*pe) + 1;
n_evtlf = n_evt_c + n_evt_e;
```

```
label sides='Sides'
      alpha='Alpha'
      power='Power'
      beta='Beta'
      zalpha='Z(alpha)'
      zbeta='Z(beta)'
      accrual='Accrual (yrs)'
      fu='Follow-up (yrs)'
      loss='Yearly Loss'
      totloss='Total Loss'
      probc='Yearly event rate: control'
      probe='Yearly event rate: active'
      rr='Hazard ratio'
      log_rr='Log(HR)'
      N='Total Sample size (RGS)'
      N_FRD='Total Sample size (Freedman)'
      N_LF='Total Sample size (L&F)'
      n_0loss='Sample size (no loss)'
      pc='Pr(event), control'
      pe='Pr(event), active'
      pc0loss='Pr(event| no loss), control'
      pe0loss='Pr(event| no loss), active'
      n_evth0='# events (Ho-pooled)'
      n_evtrgs='# events (RGS)'
      n_evtfrd='# events (Freedman)'
      n_evtlf='# events (L&F)'
      medianc='Median survival, control'
      mediane='Median survival, active'
      naive='Sample size (naive loss)';


proc print data=rgs label noobs;
title 'Sample size & expected events for comparing two survival distributions';
title2 'Using method of Rubinstein, Gail and Santer (RGS, 1981)';
title3 'Freedman (1982), or Lachin and Foulkes (L&F, 1986)';
  var sides alpha power accrual fu loss totloss
        probc probe medianc mediane pc pe pc0loss pe0loss rr log_rr
        n_evth0 n_evtrgs n_evtfrd n_evtlf N N_FRD N_LF n_0loss naive;
  format power loss totloss f4.2 medianc mediane f5.3
         probc probe rr log_rr pc pe pc0loss pe0loss f6.4;
```

Back to Example I:

A clinical trial in esophageal cancer will randomize patients to radiotherapy alone (Rx A) versus radiotherapy plus chemotherapy (Rx B). The goal of the study is to compare the two treatments with respect to survival, and we plan to use the logrank test. From historical data, we know that the median survival on Rx A for this disease is around 9 months. We want 90% power to detect an improvement in this median to 18 months. Past studies have been able to accrue approximately 50 patients per year. Choose a suitable study design.

First, let's write down what we know:

- desired significance level not stated, so use $\alpha = 0.05$ (assume a two-sided test)
- assume equal randomization to treatment arms (unless otherwise stated)
- desired power is 90%
- median survival on control is 9 months $\Rightarrow M0 = 9$
- want to detect improvement to 18 months on Rx B $\Rightarrow$ $M1 = 18$
- Maximum accrual per year is 50 patients

We have all of the information we need to run the program, except the accrual and follow up times. We need to use trial and error to get these.

Accrual Period	Follow-up Period	Number of Events Required	Total Sample Size	Total Study Duration
1	2.5	88	106	3.5
2	1.5	88	115	3.5
2.5	1	88	122	3.5
3	0.5	88	133	3.5
3	1	88	117	4

Shown on the next page is the output from `rgsnew.sas` using Accrual=2, Follow-up=1.5. I've given the RGS numbers above.

Which of the above are feasible designs?

Sample size & expected events for comparing two survival distributions
Using method of Rubinstein, Gail and Santer (RGS, 1981)
Freedman (1982), or Lachin and Foulkes (L&F, 1986)

Sides	Alpha	Power	Accrual (yrs)	Follow-up (yrs)	Yearly Loss	Total Loss	Yearly event rate: control
2	0.05	0.90	2	1.5	0.00	0.00	0.6031

Yearly event rate: active	Median survival, control	Median survival, active	Pr(event), control	Pr(event), active	Pr(event| no loss), control
0.3700	0.750	1.500	0.8860	0.6737	0.8860

Pr(event| no loss), active	Hazard ratio	Log(HR)	# events (Ho-pooled)	# events (RGS)	# events (Freedman)
0.6737	2.0000	0.6931	94	88	95

# events (L&F)	Total Sample size (RGS)	Total Sample size (Freedman)	Total Sample size (L&F)	Sample size (no loss)	Sample size (naive loss)
90	115	122	121	115	115

How do we pick from the feasible designs?

The first 4 designs all have 3 1/2 years total duration, since the follow-up period starts after the last patient has been accrued. The shorter the follow-up period given this fixed study duration, the more patients we have to enroll.

In some cases, it will be much more cost-effective to enroll fewer patients and follow them for longer. This corresponds to cases where the initial cost per patient is very high.

In other cases (where the initial cost per patient is lower), it will be better to enroll more patients. The median follow-up for the first 4 designs are 3, 2.5, 2.25, and 2 years, respectively. The total cost of treatment could be estimated by multiplying the number of patients by the median follow-up time.

Some prefer to keep the accrual period as short as possible, given how many patients can feasibly be enrolled. This will tend to give the smallest number of patients among the feasible designs. Which design would this correspond to?

Another issue to think about is whether the background conditions of the disease are changing rapidly (like AIDS) or are fairly stable (like many types of cancer). For the former situation, it would be best to have a study with a short duration so the results will have more interpretation.

Using the information given, there are a lot of other quantities we can calculate:

- The **hazard ratio of control to treated** is:

$$\frac{\text{median(Rx B)}}{\text{median(Rx A)}} = \frac{18}{9} = 2$$

- The **hazard rates** for the two treatment arms are:

$$\text{for Rx A:} \quad \lambda_0 = \frac{-\log(0.5)}{\text{median(Rx A)}} = \frac{-\log(0.5)}{9} = 0.0770$$

$$\text{for Rx B:} \quad \lambda_1 = \frac{-\log(0.5)}{\text{median(Rx B)}} = \frac{-\log(0.5)}{18} = 0.0385$$

- The **yearly probability of an event** is:

$$\text{for Rx A:} \quad Pr(T < 1|\lambda_0) = 1 - e^{(-\lambda_0 * t)}$$
$$= 1 - e^{(-0.0770*12)} = 0.603$$

$$\text{for Rx B:} \quad Pr(T < 1|\lambda_1) = 1 - e^{(-\lambda_1 * t)}$$
$$= 1 - e^{(-0.0385*12)} = 0.370$$

What would happen above if we used time t in years (i.e., t=1) instead of months?

What would happen if we calculated both the hazard rate and yearly event probability using time in years?

Based on a design with 2.5 years accrual and 1 year follow-up:

- The median follow-up time

$$\begin{aligned} \text{median FU} &= A/2 + F \\ &= 30/2 + 12 = 27 \text{ months} \end{aligned}$$

- The probability of an event during the entire study is: (using the approximation in notes)

$$\begin{aligned} \text{for Rx A:} \quad P_c &= 1 - exp(-\lambda_0 * [A/2 + F]) \\ &= 1 - exp(-0.0770 * 27) = 0.875 \end{aligned}$$

$$\begin{aligned} \text{for Rx B:} \quad P_e &= 1 - exp(-\lambda_0 * [A/2 + F]) \\ &= 1 - exp(-0.0385 * 27) = 0.646 \end{aligned}$$

(the above numbers differ from what you'd get in the printout from the program, since it calculates the exact probability under the exponential distribution, instead of using the approximation)

In the calculations above, all of the "time" periods were in terms of months. You have to remember to keep the scale the same throughout.

To use the program, you need to translate the time scale in terms of years. So a median of 18 months survival would be entered as median=1.5.

What happens if we add loss to follow-up?

Required sample size for A=2.5, FU=1 year

Yearly Loss to Follow-up	Number of Events Required	Total Sample Size	Total Study Duration
0	88	122	3.5
5%	88	128	3.5
10%	88	133	3.5
20%	88	147	3.5

Sequential Design and Analysis of survival studies

In clinical trials and other studies, it is often desirable to conduct interim analyses of a study while it is still ongoing.

Rationale:

- **ethical:** if one treatment is substantially worse than another, then it is wrong to continue to give the inferior treatment to patients.
- **timely reporting:** if the hypothesis of interest has been clearly established halfway through the study, then science and the public may benefit from early reporting.

WARNING!!
Unplanned interim analyses can seriously inflate the true type I error of a trial. If interim analyses are to be performed, it is ESSENTIAL to carefully plan these in advance, and to adjust all tests appropriately so the the type I error is of the desired size.

How does the type I error become inflated?

Consider a two group study comparing treatments A and B.

Suppose the data are normally distributed (say $X_i \sim N(\mu_A, \sigma^2)$ in group A, and similarly for group B), so that the null hypothesis of interest is

$$H_0 : \mu_A = \mu_B$$

It is not too hard to figure out how the type I error can get inflated if a naive approach is used.

Suppose we plan to do K interim analyses, and that exactly m individuals will enter <u>each</u> treatment between each analysis. The test statistic at the kth analysis will be

$$Z_k \;=\; \frac{\Sigma_{i=1}^{k}\,\Sigma_{j=1}^{m}(X_{Aij} - X_{Bij})/km}{\sqrt{2\sigma/km}} \;=\; \frac{\Sigma_{i=1}^{k}\,d_i/k}{\sqrt{2\sigma/km}}$$

where d_i is the difference between the two group means at the ith analysis,

$$d_i = \overline{X}_{Ai} - \overline{X}_{Bi}$$

and $\overline{X}_{Ai}$ and $\overline{X}_{Bi}$ are the means in groups A and B of the m individuals who entered in the i-th time period.

(naive) Interim monitoring procedure:

- Allow m patients to enter on each treatment arm (total of $2m$ additional patients)
- Calculate Z_k based on the current data
- Reject the null hypothesis if $|Z_k| > z_{1-\alpha/2}$, where α is the desired type I error.

The overall type I error rate for the study is:

$$Pr(|Z_1| > z_{1-\alpha/2} \text{ or } |Z_2| > z_{1-\alpha/2} ... \text{ or } |Z_K| > z_{1-\alpha/2})$$

If the test at each interim analysis is performed at level α, then clearly this probability will exceed α. The table below shows the Type I error rate if each test is done at $\alpha = 0.05$ for various values of K:

Number of interim analyses (K)

1	2	3	4	5	10	25
5%	8.3%	10.7%	12.6%	14.2%	19.3%	26.6%

(from Lee, *Statistical Methods for Survival Data*, Table 12.9)

For survival data, the calculations become MUCH more complicated since the data collected within each time interval continues to change as time goes on!

What can we do to protect against this type I error inflation?

Pocock Approach:

Pick a smaller significance level (say α') to use at each interim analysis so that the overall type I error stays at level α.

A problem with the Pocock method is that even the very last analysis has to be performed at level α'. This tends to be very conservative at the final analysis.

O'Brien and Fleming Approach:

A preferable approach would be to vary the alpha levels used for each of the K interim analyses, and try to keep the very last one "close" to the desired overall significance level. The O'Brien-Fleming approach does that.

Comments and notes:

- There are several other approaches available for sequential design and analysis. The **O'Brien and Fleming** approach is probably the most popular in practice.

- There are many variations on the theme of sequential design. The type we have discussed here is called **Group sequential analysis**.
  - There are other approaches that require continuous analysis after each new individual enters the study!
  - There are also approaches where the randomization itself is modified as the trial proceeds. E.g. Zelen's "Play the winner rule" (New England Journal of Medicine 300, 1979, page 1242) and Ware's "ECMO" study (Statistical Science, 4, 1989, page 298)

- Some designs allow for early stopping in the absence of a sufficient treatment effect as the trial progresses. These procedures are referred to as "stochastic curtailment" or "conditional power" calculations.

- Designing a group sequential trial for survival data requires sophisticated and highly specialized software. EaSt, a package from CYTEL SOFTWARE that does standard (fixed) survival designs, as well as sequential designs.

- Many "non-statistical" issues enter decisions about whether or not to stop a trial early

- P-values based on analyses of studies with sequential designs are difficult to interpret.

- Once you do 5 interim analyses, then adding more makes little difference. Some clinical trials groups (HSPH AIDS group) have large randomized Phase III studies monitored at least once per year (for safety reasons), and most studies have 1-3 interim looks.

- Going from a fixed to a group sequential design adds only about 3-4% to the required maximum sample size. This is a good rule of thumb to use in calculating the sample size when you plan on doing interim monitoring.

Competing Risks and Multiple Failure Times

So far, we've been acting as if there was only one endpoint of interest, and that censoring due to death (or some other event) was independent of the event of interest.

However, in many contexts it is likely that the time to censoring is somehow correlated with the time to the event of interest. In general, we often have several different types of failure (death, relapse, opportunistic infection, etc) which are related (i.e., dependent or "competing" risks).

Examples:

- After a bone marrow transplantation, patients are followed to evaluate "*leukemia-free survival*", so the endpoint is time to leukemia relapse or death. This endpoint conists of two types of failures (competing risks):
  - leukemia relapse
  - non-relapse deaths
- In cardiovascular studies, deaths from other causes (such as cancer) are considered competing risks.
- In actuarial analyses, we look at time to death, but want to provide separate estimates of hazards for each cause of death (multiple decrement lifetables).

Another example: For the MAC study, the analyses you have been doing of time to MAC assume that the censoring time is independent.

Recall:

$$\begin{aligned} T &= \text{time to event of interest (MAC)} \\ U &= \text{time to censoring (death, loss to FU)} \\ X &= min(T, U) \\ \delta &= \mathbf{I}(T \leq U) \end{aligned}$$

Observable Data: (X, δ)

What are the possiblities here?

- (1) Failure T and censoring U are independent
- (2) Failure T and censoring U are dependent

Case (1): Independent failure times
(this includes the case of independent censoring)

BOTTOM LINE $\Rightarrow$ NO PROBLEM

Nonparametric estimation:

In this case, we can use the Kaplan-Meier estimator to estimate $S_T(t) = P(T > t)$.

Parametric estimation:

If we know the joint distribution of (T, U) has a certain parametric form (exponential, Weibull, log-logistic), then we can use the likelihood for (X, δ) to get parameter estimates of the marginal distribution of $S_T(t)$.

Semi-parametric estimation:

We can apply the Cox regression model to assess the effects of covariates on the marginal hazard.

Case (2): Dependent failure times

BOTTOM LINE ⇒ BIG PROBLEM

Tsiatis (1975) showed that $S_T(t) = P(T \geq t)$ (i.e., the survival function for the event T of interest) cannot be "identified" from data of the form (X, δ) for each subject.

In fact, observing (X, δ) does not provide enough information to estimate the joint distribution of (T, U) so that we can even check whether the assumption of independence is valid.

When is it reasonable to assume independent risks?

- when censoring occurs because the study ends, or because the subject moves to a different state
- and there is no trend over time in health status of enrolling patients

In the case of our MAC study, the fact that someone dies may reflect that they would have been at greater risk of MAC if they had not died than someone else who remained alive at that point.

The assumption of independence means that the hazard for someone who is censored at time t is exactly the same as that for someone with the same covariates who is also at risk at time t.

What is the impact of dependent competing risks?

Slud and Byar (1988) show that dependent causes of death can potentially make risk factors appear protective:

If we have

$$\begin{aligned} T &= \text{death from cause of interest} \\ \text{and} \quad U &= \text{censoring, from death due to other cause} \end{aligned}$$

and a single binary covariate Z

$$Z = \begin{cases} 1 & \text{if risk factor is present} \\ 0 & \text{otherwise} \end{cases}$$

and we calculate the Kaplan-Meier survival estimates $\hat{S}_1(t)$ for $Z = 1$ and $\hat{S}_0(t)$ for $Z = 0$ assuming independent censoring, then we could (in their hypothetical example) end up reversing the sign of the survival functions:

True ordering between survival distributions:

$$S_1(t) < S_0(t) \qquad \text{for all } t$$

Kaplan Meier estimates of survival distributions:

$$\hat{S}_1(t) > \hat{S}_0(t) \qquad \text{for all } t$$

What can we do if we suspect dependent risks?

A lot of people have tried to tackle this problem!

References

Aly EAA, Kochar SC, and McKeague IW (1994). Some tests for comparing cumulative incidence functions and cause-specific hazard rates. *JASA* **89**, 994-999.

Benichou J and Gail MH (1990). Estimates of absolute cause specific risk in cohort studies. *Biometrics* **46**, 813-826.

(*) Booth AL and Satchell SE (1995). The hazards of doing a PhD: an analysis of completion and withdrawal rates of British PhD students in the 1980's. *JRSS-A*, 297-318.

Farewell VT (1979). An application of Cox's proportional hazard model to multiple infection data. *Applied Statistics* **28**, 136-143.

Gail, M (1982). Competing risks. *Encyclopedia of Statistical Sciences* **2**, 75-81.

Lin DY, Robins JM, and Wei LJ (1996). Comparing two failure time distributions in the presence of dependent censoring. *Biometrika* **83**, 381-393.

Lunn M and McNeil D (1995). Applying Cox regression to competing risks. *Biometrics* **51**, 524-532

Moeschberger ML and Klein JP (1988). Bounds on net survival probabilities for dependent competing risks. *Biometrics* **44**, 529-538.

(*) Moeschberger ML and Klein JP (1995). Statistical methods for dependent competing risks. *Lifetime Data Analysis* **1**, 193-204.

Pepe MS (1991). Inference for events with dependent risks in multiple endpoint studies. *JASA* **86**, 770-778.

(*) Pepe MS and Mori M (1993). Kaplan-Meier, marginal or conditional probability curves in summarizing competing risks failure time data? *Statistics in Medicine* **12**, 737-751.

Prentice RL, Kalbfleisch JD, Peterson AV, Flournoy N, Farewell VT, and Breslow NE (1978). The analysis of failure times in the presence of competing risks. *Biometrics* **34**, 541-554.

Slud EV, Byar DP, and Schatzkin A (1988). Dependent competing risks and the latent failure model. *Biometrics* **44**, 1203-1205.

Slud E and Byar D (1988). How dependent causes of death can make risk factors appear protective. *Biometrics* **44**, 265-269.

Tsiatis, A. (1975). A nonidentifiability aspect of the problem of competing risks. *Proceedings of the National Academy of Sciences* **72**, 20-22.

There has been a lively debate in the literature about the best way to attack this problem. The two sides are basically divided about which type of model to use:

- based on **cause-specific hazard** functions (observables)
- based on **latent variable** models (unobservables)

The first approach focuses on what the observed survival is due to a certain cause of failure, acknowleding that there are other types of failures operating at the same time.

The second approach attempts to estimate what the survival associated with a certain failure type *would* have been, if the other types of failures had been removed.

General Case of Multiple Failure Types

In general, say we have m different types of failure (say, causes of death), and the respective times to failure are:

$$T_1, T_2, T_3, \cdots, T_m$$

and we observe $T = min(T_1, T_2, \ldots, T_m)$

We can write the *cause-specific hazard function* for the j-th failure type as:

$$\lambda_j(t) \;=\; \lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr(t \leq T < t + \Delta t, J = j | T \geq t)$$

The overall hazard of death is the sum over the failure types:

$$\lambda(t) \;=\; \sum_{j=1}^{m} \lambda_j(t)$$

$$\text{where} \quad \lambda(t) \;=\; \lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr(t \leq T < t + \Delta t | T \geq t)$$

Q. Can we estimate these quantities? ...even if the risks are dependent?

A. Yes, Prentice (1978) shows that probabilities that can be expressed as a function of the cause-specific hazards can be estimated.

For example, estimable quantities include:

(a) **The overall survival probability**[3]**:**

$$\begin{aligned} S_T(t) = P(T \geq t) &= \exp\left[-\int_0^t \lambda(u)du\right] \\ &= \exp\left[-\int_0^t \sum_j \lambda_j(u)du\right] \end{aligned}$$

(b) **Conditional probability of failing from cause** j in the interval $(\tau_{i-1}, \tau_i]$

$$Q(i,j) = [S_T(\tau_{i-1})]^{-1} \int_{\tau-1}^{\tau} \lambda_j(u)\, S_T(u)\, du$$

(c) **Conditional probability of surviving** i^{th} **interval**

$$\rho_i = 1 - \sum_{j=1}^{m} Q(i,j)$$

[3]Note: previously I said you couldn't estimate $S_T(t)$, but that was when T was the time to event of interest (possibly unobservable) and U was the possibly correlated censoring time. Here, $S_T(t)$ is the survival distribution for the minimum of all failures, which can always be observed

Estimators:

(a) **The MLE of $Q(i,j)$ is simply**

$$\hat{Q}(i,j) \;=\; \frac{d_{ij}}{r_i}$$

i.e., the number of failures (deaths) due to cause j during the i-th interval among the r_i subjects at risk of failure at the beginning of the interval.

(b) **The MLE of ρ_i** is::

$$\hat{\rho}_i \;=\; \frac{r_i - \Sigma_{j=1}^{m} d_{ij}}{r_i} \;=\; 1 - \frac{\Sigma_{j=1}^{m} d_{ij}}{r_i}$$

(c) **The MLE of $S_T(t)$** is based on ρ_i:

$$\hat{S}_T(\tau_i) \;=\; \prod_{k=1}^{i} \rho_k$$

So what can't we estimate?

Compare the *cause-specific hazard function*:

$$\lambda_j(t) \;=\; \lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr(t \leq T < t + \Delta t, J = j | T \geq t)$$

with the *marginal hazard function*:

$$\lambda_j(t) \;=\; \lim_{\Delta t \to 0} \frac{1}{\Delta t} Pr(t \leq T_j < t + \Delta t | T_j \geq t)$$

We can get estimates of the cause-specific hazard function, since we can estimate $S_T(t) = P(T \geq t)$ even if the failure times are dependent. (In other words, we can observe whether each patient is still alive or not)

But unfortunately, we can't estimate the marginal hazard function when the risks are dependent, since we can't estimate $S_j(t) = P(T_j \geq t)$. (we can't tell when they would have had event T_j if they have a different event first)

This is the main tricky issue of competing risks analyses.

Back to original question...

What can we do if we suspect dependent risks?

Ex. Say we have two types of failures, T_1 and T_2, and we think they are dependent. However, we are interested in the first type of failure T_1, and view the competing risk T_2 as censoring (like in bone marrow transplant example).

Methods of summarizing data with competing risks:

(1) Summarize the **cause-specific hazard** rate over time

(2) Use the **Kaplan-Meier** estimate anyway, $\hat{S}_{T_1}(t)$

(3) Report the **complement** to the KM, $1 - \hat{S}_{T_1}(t)$

(4) Use **cumulative incidence** curves (crude incidence curve)

(5) Use the **conditional probability** function

(6) Give upper and lower **bounds for the true marginal** survival function, in the absence of the competing risk

Pepe and Mori review the first 5 of these options, and recommend *against* option (2), but note that this is often what people end up doing.

To make the example more concrete:

Say we are interested in time to MAC or death, whichever occurs first. We define:

$$\begin{aligned} T &= \text{time to MAC or death} \\ \text{and} \quad U &= \text{censoring (assumed independent)} \end{aligned}$$

and the type of failure is denoted by j

$$j = \begin{cases} M & \text{if event is MAC} \\ D & \text{if event is death from other causes} \end{cases}$$

In the alternative "latent variable" framework, we would define

$$\begin{aligned} T_M &= \text{time to MAC} \\ \text{and} \quad T_D &= \text{time to Death} \end{aligned}$$

although we might not be able to observe T_M if T_D occurred first.

Methods for competing risks:
(1) Summarizing the cause-specific hazard over time

As mentioned above, this is one of the quantities that we can estimate. Our basic estimator during time interval i is $\hat{Q}(i,j) = \frac{d_{ij}}{r_i}$.

So we can plot $\hat{\lambda}_j(t)$ over time, and get some insight as to biological phenomona involved.

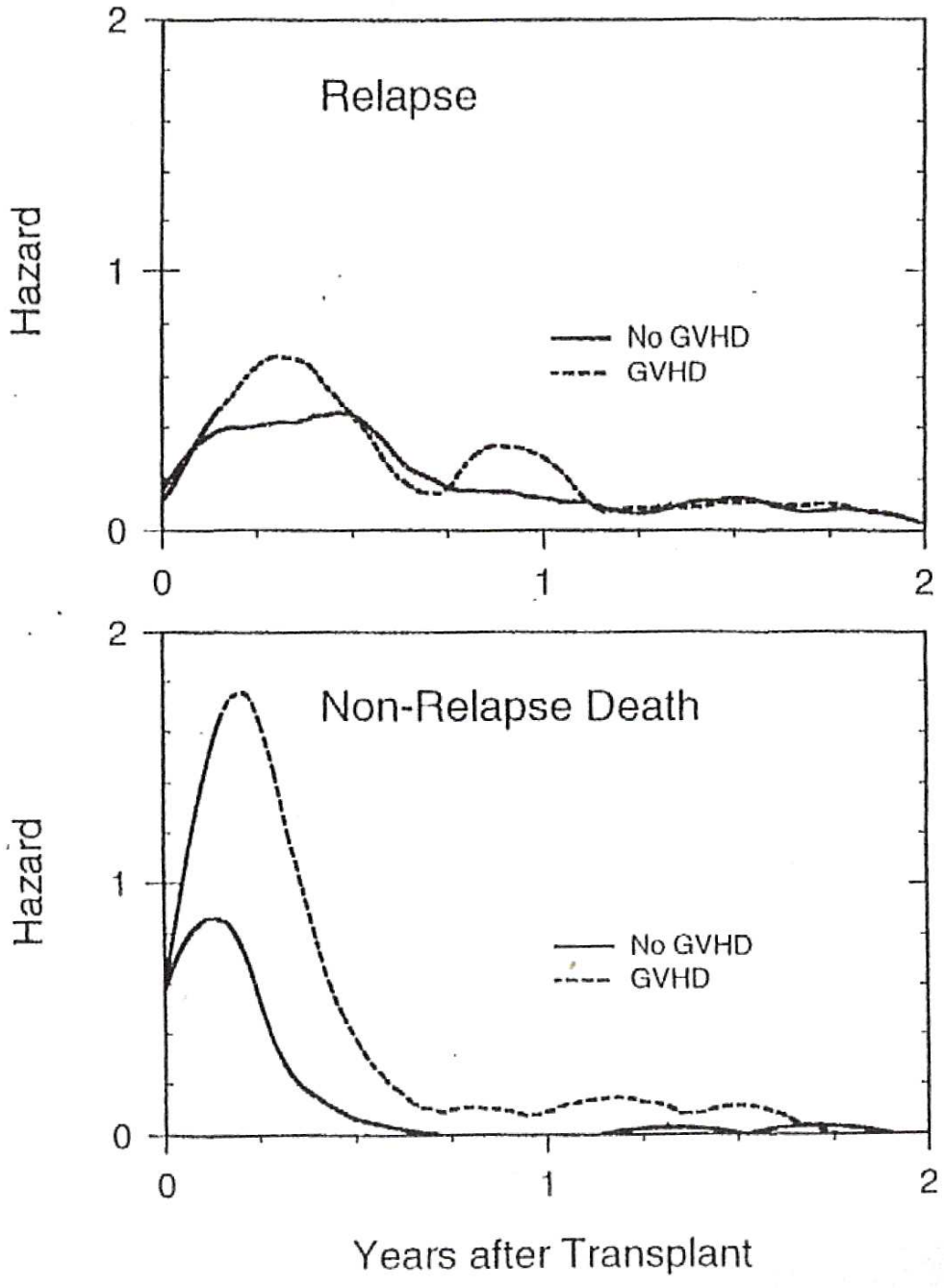
However, if you remember some of the plots I showed you of hazards over time, they tended to be highly variable. Several contributions have been made towards "smoothing" out the inherent variability in the estimates of cause-specific hazards.

- Efron (1988)
- Ramlau-Hansen (1983)
- Tanner and Wong (1983)

Drawback: the hazard functions alone do not give overall effect of a covariate on survival.

Example: If the hazard functions for $\lambda_j(t)$ for two treatments cross, we can't say which treatment leads to lower overall event rate.

Figure from Pepe and Mori for Leukemia Data
Kernel Estimates of Cause-Specific Hazards



Methods for competing risks:
(2) Applying Kaplan-Meier to cause-specific λ_j's

Say we evaluate the MAC survival distribution by treating

- all MAC cases as "events"
- any deaths without MAC as "censorings"

and construct the Kaplan-Meier survival curve.

What are we estimating?

$$S^*_M(t) \;=\; \exp\left[-\int_0^t \lambda_M(u)\; du\right]$$

where $\lambda_M(t)$ is the cause-specific hazard for MAC:

$$\lambda_M(t) \;=\; \lim_{\Delta t \to 0} \frac{1}{\Delta t}\, Pr(t \le T < t + \Delta t, j = M | T \ge t)$$

i.e., the conditional probability that MAC occurs in a short period of time, given that the subject is *alive* and *MAC-free*.

The interpretation of the Kaplan-Meier curve is as the

"**exponential of the negative cumulative cause-specific hazard**".

Clinicians (and others!) have difficulty understanding this function, since it has no direct clinical interpretation.

Figure from Pepe and Mori for Leukemia Data Kaplan Meier with Cause Specific Hazards

They thought this was such a bad idea, that they did not include any plot of this!

Methods for competing risks:
(3) Using the complement of the Kaplan-Meier

Another function used fairly often in the competing risks area is sometimes referred to as **the pure probability function**:

$$1 - S_j^*(t) \;=\; 1 - \exp\left[-\int_0^t \lambda_j(u)\, du\right]$$

In our MAC example, $1 - S_M^*(t)$ could be interprested as the predictive probability of MAC by time t if the risk of death could be removed.

If we were designing a new study for a miracle drug that seemed so powerful that it would not only reduce MAC but prevent all death from other causes in HIV-infected patients, then we could use estimates $1 - \hat{S}_M^*(t)$ to help design our new study.

This would be pretty optimistic, and there has been a lot of work on the strict (and *untestable*) assumptions required to interpret the KM curve in this manner.

Pepe and Mori contend that this function is **irrelevant** for summarizing data from a competing risks study.

Figure from Pepe and Mori for Leukemia Data Complement Kaplan-Meier Functions

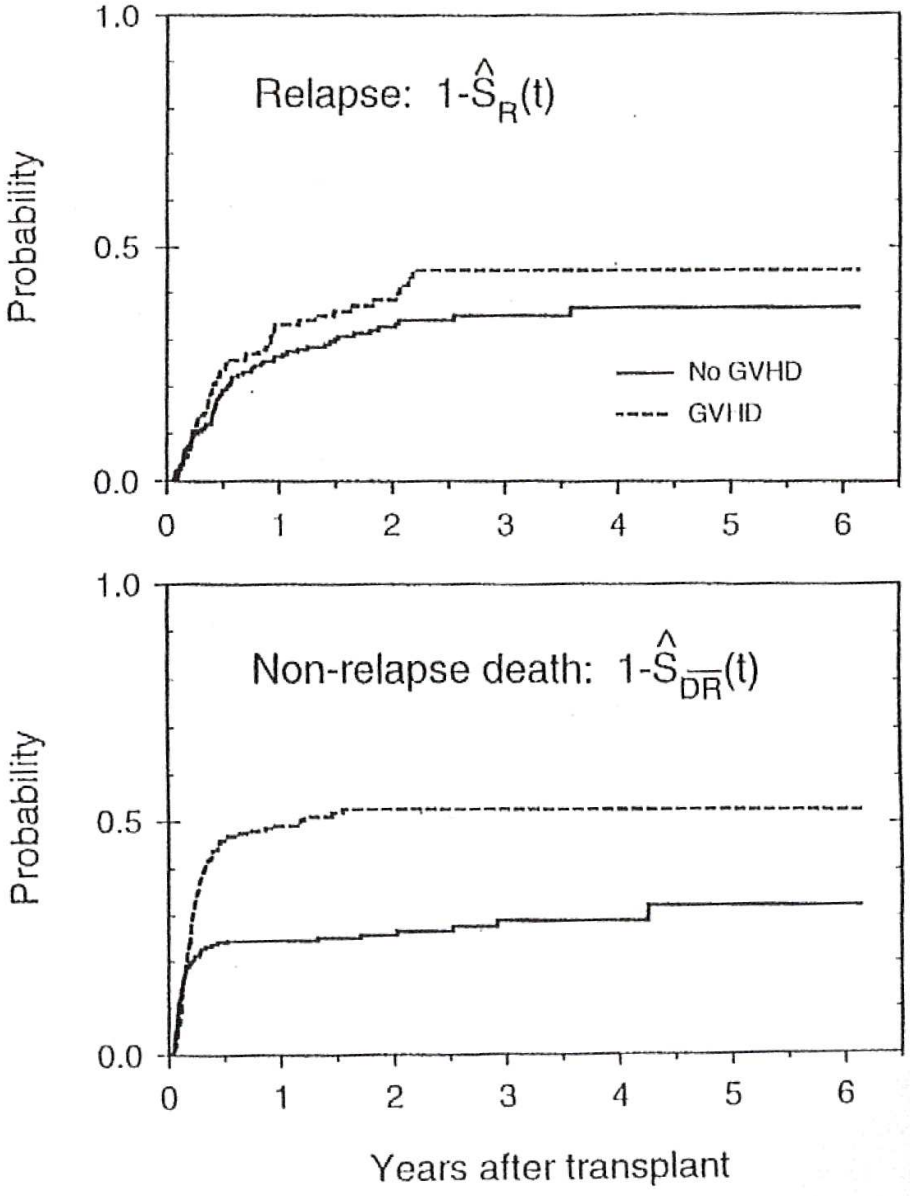


Figure 1. Complement Kaplan–Meier functions

Methods for competing risks:
(4) Using Cumulative Incidence Curves

This has also been termed the "crude incidence curve", and estimates the **marginal probability of an event** in the setting where other competing risks are acknowledged to exist.

Description of Method: the method is described in more detail in Kalbfleisch and Prentice (p.169).

Tests for covariates: Tests for comparing cumulative incidence curves among treatment groups (or some other covariate) have been developed by Bob Gray (1988). They are similar to logrank tests in that they are "linear rank" statistics.

If we were able to follow up all subjects to time t, then the cumulative incidence curves would relfect what proportion of the total study population have had the particular event (i.e., MAC) by time t.

Figure from Pepe and Mori for Leukemia Data Cumulative Incidence Curves

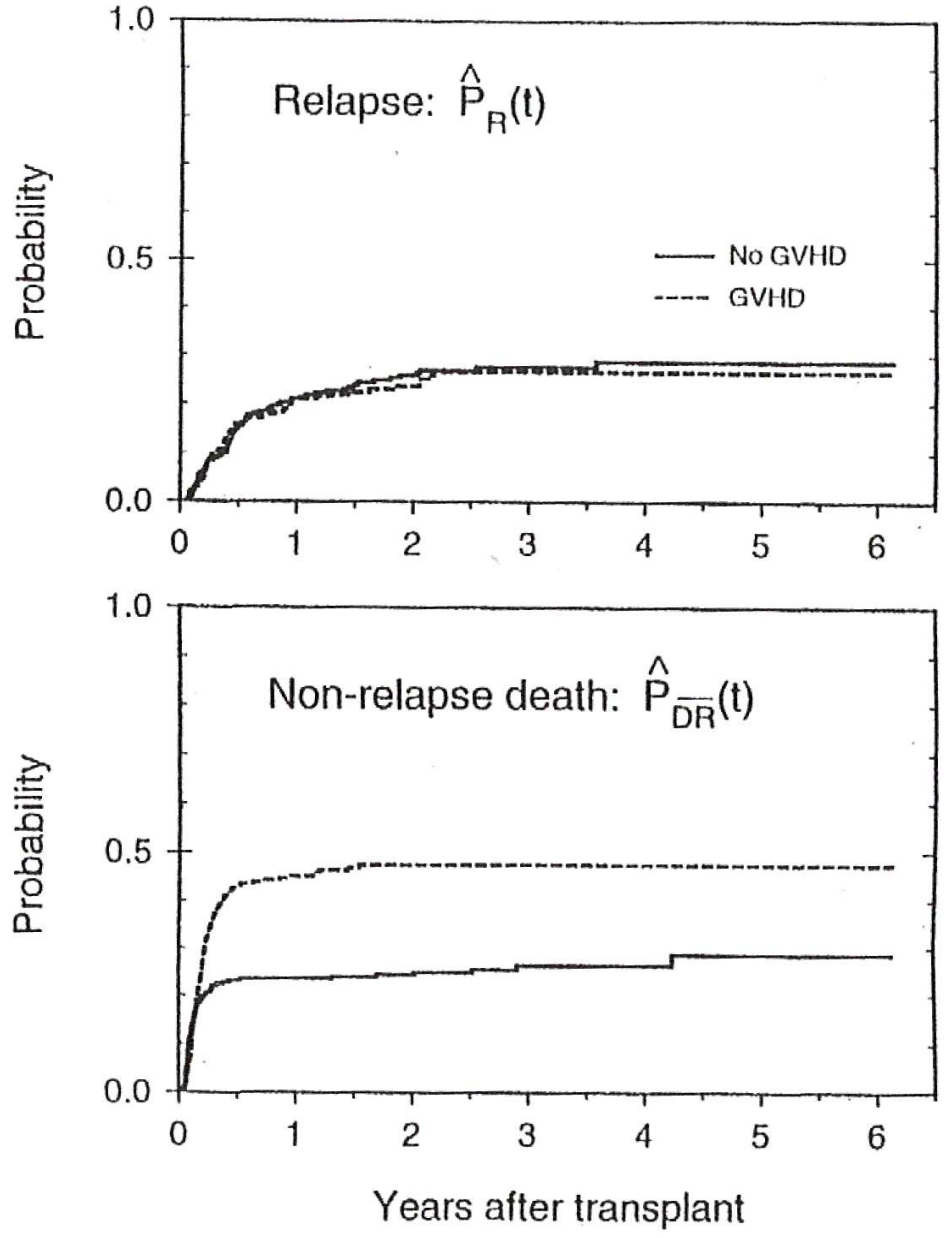


Figure 3. Marginal probability functions. Cumulative incidence estimates were used

Methods for competing risks:
(5) Conditional Probability Curves

This has the same flavor as the complement KM, but a more natural interpretation. Pepe and Mori define the conditional probability function as:

$$\begin{aligned} \widehat{CP}_M(t) &= P(T_M \leq t | T_D \geq t) \\ &= \frac{\hat{P}_M(t)}{1 - \hat{P}_D(t)} \end{aligned}$$

$$\text{where} \quad \hat{P}_M(t) = \int_0^t \hat{S}_T(u) \, \frac{d\, N_M(u)}{Y(u)}$$

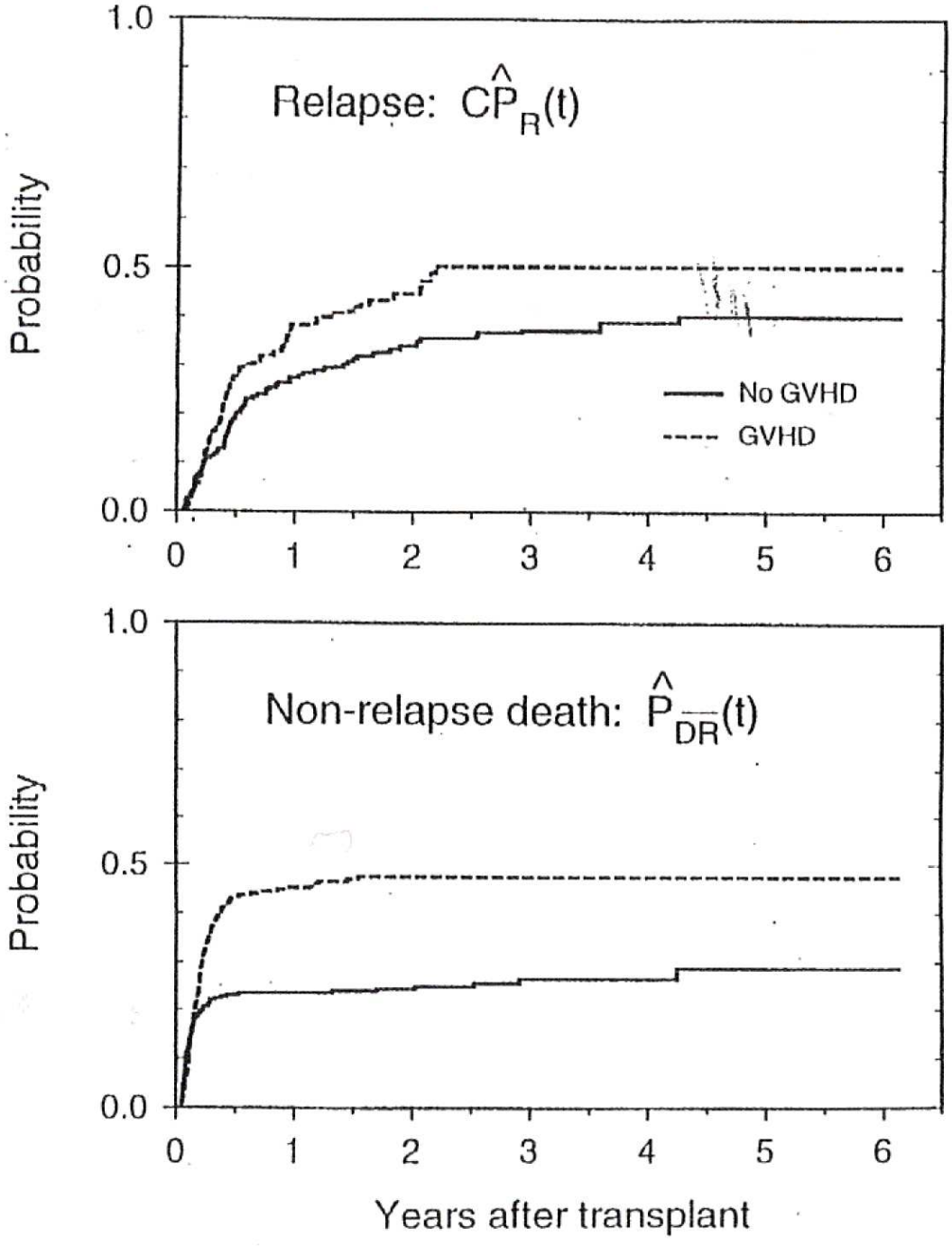
$$\text{and} \quad \hat{P}_D(t) = \int_0^t \hat{S}_T(u) \, \frac{d\, N_D(u)}{Y(u)}$$

In the above, $\hat{S}_T(u)$ is the KM estimate of the overall mac-free survival distribution, and the terms $N_M(u)$, $N_D(u)$, and $Y(u)$ reflect the "counting process" for the number of subjects with MAC, death, and at risk at time u, respectively.

In the absence of censoring, the interpretation is the proportion of patients who develop MAC among those who do not die of other causes.

Tests: Pepe and Mori also present tests, which are sums over time of weighted differences between the conditional (or marginal) probabilities for two groups.

Figure from Pepe and Mori for Leukemia Data
Conditional Probability and Marginal Curves



Methods for competing risks:
(6) Bounds on Net Survival Curves

As noted previously, we cannot estimate $S_j(t) = P(T_j \geq t)$ if the failure times are dependent (eg, we can't estimate the survival function for MAC if time to MAC is correlated with time to death without MAC).

However, we may be able to say something about the range of $S_j(t)$ by finding upper and lower bounds that contain $S_j(t)$.

- **Peterson** (1976) obtained general bounds based on the minimal and maximal dependence structure for (T_M, T_D). The bounds allow any possible dependence structure, but can be *very* wide.
- **Slud and Rubenstein** (1983) obtained tighter bounds on $S_j(t)$ by using additional information, but require the user to specify reasonable bounds on a function ρ. Once ρ is supplied, the marginal distribution $S_{j,\rho}(t)$ can be obtained.
- **Klein and Moeshberger** (1988) use the framework of Clayton and Oakes for bivariate survival to obtain tighter bounds than those of Peterson. Again, the user has to supply bounds on a function θ, and once this is given $\hat{S}_{j,\theta}(t)$ can be obtained.

Figure from Klein and Moeschberger (1988) Bounds on Net Survival Curves

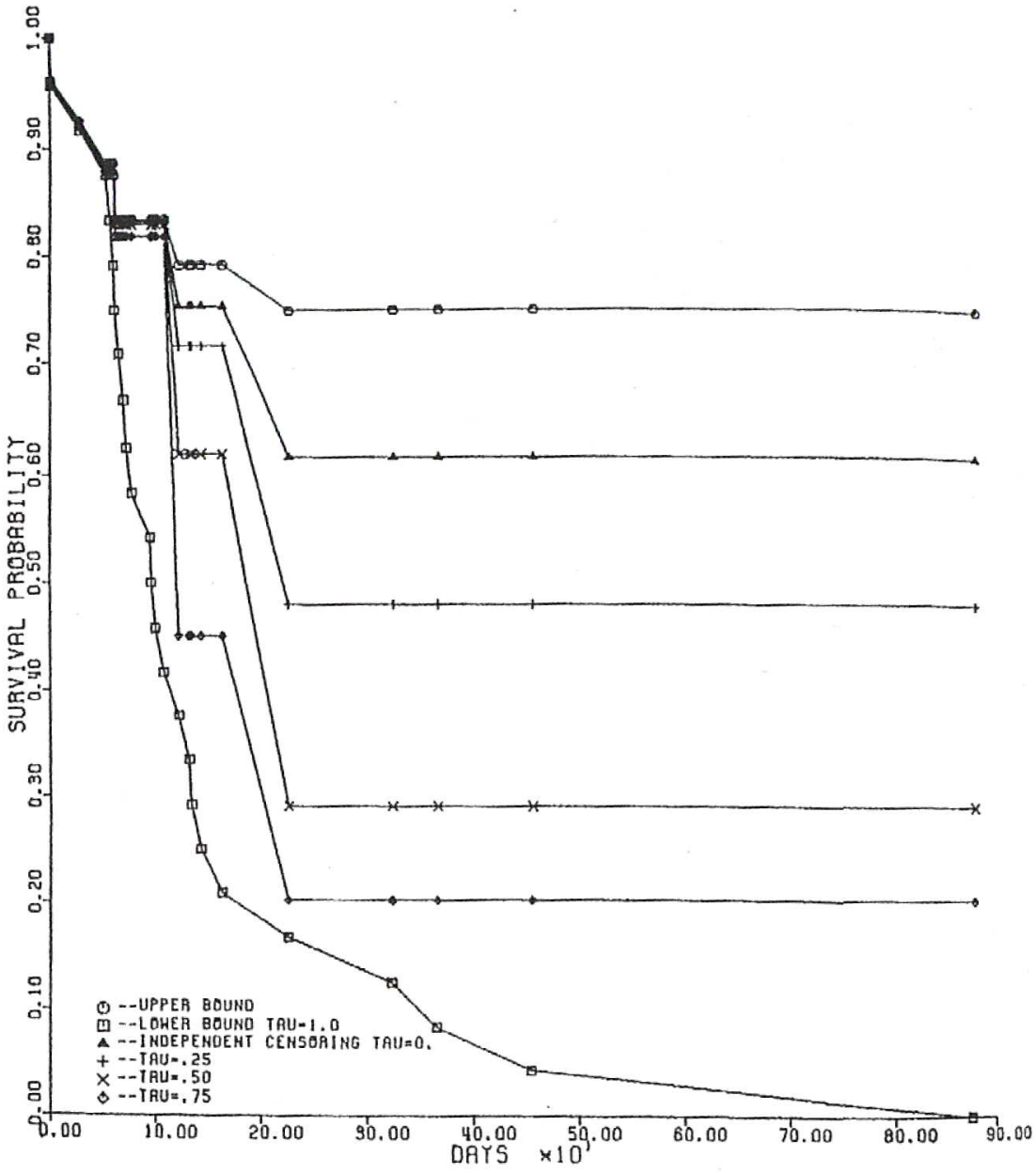


Figure 2. Bounds on survival for non-Hodgkins lymphoma patients.

One last example: Promotion of Faculty at HSPH

I was asked to analyze the school's data from 1980-1995 on promotion of Faculty from Assistant Professor to Associate Professor to assess whether there were differences between males and females and among academic areas (social, laboratory, quantitative).

Problem:

Would you think that "censoring" (someone leaving their tenure track position prior to getting promoted) is independent of the probability of promotion?

I considered 3 approaches for accounting for censoring:

```
Method I:   assumes those who departed would NOT have
            been promoted

Method II:  assumes those who departed would have been
            promoted at the same rate as those who stayed

Method III: assumes 50% of those who departed would not
            have been promoted, and the other 50% would
            have been promoted at the same rate as those
            who stayed
```

Which Method corresponds to "non-informative" (independent) censoring?

Results: Cumulative probabilities of promotion

Effect of Gender on Promotion

	Overall	Males	Females
Method I:	0.631	0.719	0.451
Method II:	0.933	1.000	0.674
Method III:	0.736	0.825	0.531

Effect of Academic Area on Promotion

	Overall	Quantitative	Social	Laboratory
Method I:	0.631	0.703	0.238	0.701
Method II:	0.933	0.950	0.389	1.000
Method III:	0.736	0.803	0.287	0.801