1) p1 This course considers the analysis of "time to event" data.
Engineers use reliability analysis studies the lifetime (time until failure) of manufactured products.

Survival analysis studies the lifetime (time until death) of humans, often after contracting a deadly disease.
In the social sciences, the study of the time until the occurence of an event, eg 1st use of cigarettes or time of 1st morriage, is called the analysis of event time data or event history analysis.
In economics, the study is called duration analysis or transition analysis.

2) Actuaries put prices on risks, so survival analysis is an important topic for actuarial sciences, statistics and biostatistics

3) survival data, reliability data, failuretime data, lifetime data and event time data are very similar.

4) p. 3 The survival time of an individual is censored if the event of interest has not been observed.
- censoring occurs because of cost and time constraints.

ex) study 100 breast cancer patients who receive a lumpectomy for 5 years.

Perhaps 15% die 5% move and 80% are still alive at the end of 5 years.

Then 85% of the 61 cases are censored.

5) Let Y_i = time until event (eg death) for ith person.

Let Z_i = time ith person leaves study for any reason other than the event of interest = time until ith person is censored.

Then the (right censored) survival time

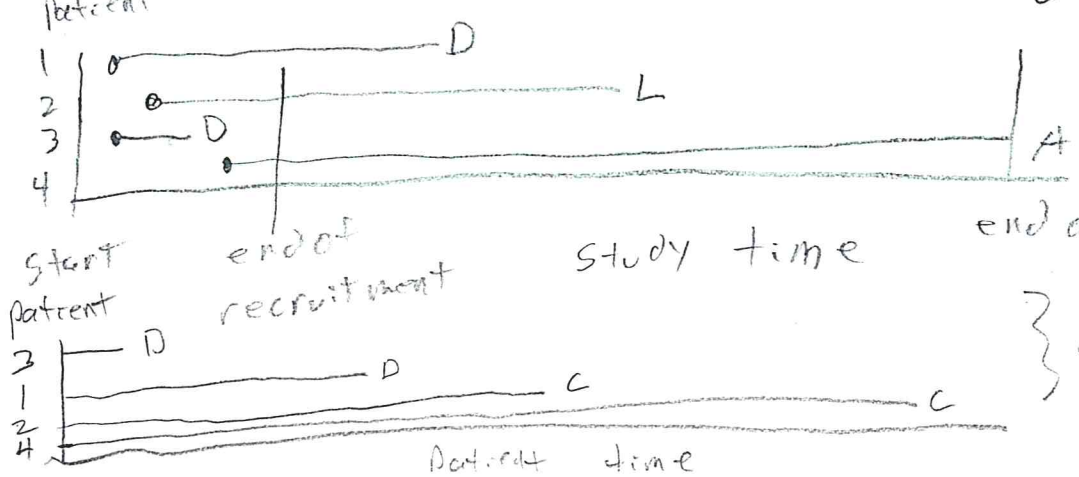
$T_i = \min(Y_i, Z_i)$.

Let $\delta_i = \begin{cases} 0 & \text{if } T_i \text{ is censored } (T_i = Z_i) \\ 1 & \text{if } T_i \text{ is not censored } (T_i = Y_i). \end{cases}$

Data: $(T_1, \delta_1), \dots, (T_n, \delta_n)$ (eg for computer file)

or $T_1, T_2, T_3^*, \dots, T_{n-1}^*, T_n$ where * means censored

ex) p. 19 D = died L = lost to study A = alive at end of study C = right censored



6/7 Assume T_i and C_i are independent. The censoring mechanism is independent of the time to event.

Skim §1.2 for examples

§1.3 7) know for final p27, 73, 81, 82 Let T be a nonnegative random variable, eg time until death. Each of the following 5 quantities completely determines the distribution of T.

A) The cumulative distribution function (cdf) $F(t) = P(T \le t)$.

B) The probability density function (pdf) text says $P(T<t)$

$$f(t) = \frac{d\,F(t)}{dt} = F'(t).$$

C) The survival function $S(t) = P(T > t) = 1 - F(t)$

D) The hazard function $h(t) = \frac{f(t)}{1-F(t)}$, $t > 0$.

E) The cumulative hazard function $H(t) = \int_0^t h(u)\,du$, $t \ge 0$.

Since $\frac{d}{du}[-\log(1-F(u))] = \frac{-(-1)(-F'(u))}{1-F(u)} = \frac{f(u)}{1-F(u)}$, $H(t) = \int_0^t \frac{f(u)}{1-F(u)}\,du = -\log(1-F(u))\Big|_0^t = -\log(1-F(t)) + 0 = -\log(S(t))$

Hence $S(t) = e^{-H(t)}$, Note $S(0) = 1$ so $H(0) = 0$. $S(\infty) = 0 \Rightarrow H(\infty) = \infty$.

8) Know $0 \le F(t) \le 1$, $0 \le S(t) \le 1$, $f(t) \ge 0$, $h(t) \ge 0$, $H(t) \ge 0$,

In actuarial sciences, $h(t)$ = force of mortality.

9) In reliability analysis the reliability function $R(t) = S(t)$. In economics, mills ratio $= \frac{1}{h(t)}$.

See p8 from E1 rev $S(t) = \exp(-H(t)) \Rightarrow H(t) = -\log S(t)$

10) p11-12 Note that $P[t < T \le t + \Delta t \mid T > t]$

$A \subseteq B$ so $A \cap B = A$

$$= \frac{P[t < T \le t + \Delta t]}{P(T > t)} = \frac{F(t+\Delta t) - F(t)}{1 - F(t)}, \text{ so}$$

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

$$\lim_{\Delta t \to 0} \frac{1}{\Delta t} P(t < T \le t+\Delta t \mid T > t) = \lim_{\Delta t \to 0} \frac{\left(\frac{F(t+\Delta t) - F(t)}{\Delta t}\right)}{1 - F(t)}$$

$$= \frac{f(t)}{1-F(t)} = h(t).$$ So for small Δt (eg. 10 of years or 1 day)

$h(t)\Delta t \approx P(t < T \le t+\Delta t \mid T > t) \approx P$(person dies in interval $(t, t+\Delta t]$ given that the person has survived up to time t). Larger $h(t)$ means hazard of death is higher. The hazard function takes into account <u>aging</u> (of a person or product).

ex) An 80 year old male has about a 50% chance of living to 85. A 100 year old person has about a 50% chance of living to 101. Hardly any people will live to 101.

11) Since $T \ge 0$, $F(0) = 0$, $F(\infty) = 1$ F is non decreasing.

Similarly $S(0) = 1$, $S(\infty) = 0$ and S is nonincreasing.

we typically graph $S(t)$ for $t \ge 0$. ($S(t) = 1$ for $t < 0$)

12) <u>know for real</u> Given one of $F(t)$, $f(t)$, $S(t)$, $h(t)$ or $H(t)$, know how to find the other 4 quantities for $t > 0$.

A) If $f(t)$ is given, $F(t) = \int_0^t f(u)du$, $S(t) = \int_t^\infty f(u)du = 1-F$

$h(t) = \frac{f(t)}{1-F(t)} = \frac{f(t)}{S(t)}$, $H(t) = \int_0^t h(u)du = -\log[S(t)]$.

B) If $F(t)$ is given, $f(t) = F'(t)$, $S(t) = 1 - F(t)$,

$h(t) = \frac{f(t)}{1-F(t)} = \frac{f(t)}{S(t)}$, $H(t) = \int_0^t h(u)du = -\log(S(t))$.

C) If $S(t)$ is given, $F(t) = 1 - S(t)$,

$f(t) = h(t)S(t) = F'(t) = -S'(t)$, $h(t) = \frac{f(t)}{S(t)} = \frac{-S'(t)}{S(t)}$,

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

D) If $h(t)$ and $H(t)$ are given,

$F(t) = 1 - \exp\left[-\int_0^t h(u)du\right] = 1 - e^{-H(t)}$.

$f(t) = h(t)\exp\left[-\int_0^t h(u)du\right] = h(t)e^{-H(t)} = H'(t)e^{-H(t)} = \lambda e^{-\lambda t}$

$S(t) = \exp\left[-\int_0^t h(u)du\right] = e^{-H(t)}$.

Note that $h(t) = H'(t)$. Hence each of the 4 functions is a "simple" function of $H(t)$.

Some Parametric Distributions (see Olive & Ibl)

1] ex] The exponential distibution $T \sim EXP(\lambda)$ has $h(t) = \lambda$ for $t > 0, \lambda > 0$. Find $F(t)$, $f(t)$, $S(t)$ and $H(t)$.

Soln] $H(t) = \int_0^t h(u)du = \int_0^t \lambda\, du = \lambda t$, $t > 0$.

$S(t) = e^{-H(t)} = e^{-\lambda t}$, $t > 0$

$F(t) = 1 - S(t) = 1 - e^{-\lambda t}$, $t > 0$

$f(t) = h(t)S(t) = \lambda e^{-\lambda t} = F'(t)$

2] If $T \sim EXP(\lambda)$, then $E(T) = \frac{1}{\lambda}$.

Since $h(t) = \lambda$ is constant, risk of failure is the same for used products as for new products. So aging has no effect on the product. $T \sim EXP(\lambda)$ can

be a good model if failures are due to random shocks that follow a Poisson process. Some electrical components, eg light bulbs, are fit fairly well by the exponential distribution.
memoryless property $P(T>t) = P(T>t+a \mid T>a)$, $t, a>0$

2] p 154 T has a Weibull (λ, γ) distribution if

$$h(t) = \lambda \gamma t^{\gamma-1} \quad \text{for } t>0,\ \lambda>0,\ \gamma>0.$$

This distribution is the most important distribution in reliability analysis. The hazard function can be increasing, decreasing or constant. The Weibull $(\lambda, \gamma=1)$ distribution is the $Exp(\lambda)$ distribution

ex] Find $F(t)$, $f(t)$, $S(t)$ and $H(t)$ if $T \sim$ Weibull (λ, γ).

Solution] $H(t) = \int_0^t h(u)\,du = \int_0^t \lambda\gamma u^{\gamma-1}\,du$

$$= \lambda\gamma \left.\frac{u^\gamma}{\gamma}\right|_0^t = \lambda t^\gamma, \quad t>0$$

$$S(t) = \exp(-H(t)) = \exp(-\lambda t^\gamma), \quad t>0.$$

$$F(t) = 1 - S(t) = 1 - \exp(-\lambda t^\gamma), \quad t>0$$

$$f(t) = h(t)S(t) = \lambda\gamma t^{\gamma-1}\exp(-\lambda t^\gamma), \quad t>0.$$

3] Recall from the central limit theorem that $\bar{Y} = \frac{1}{n}\sum_{i=1}^{n} Y_i \approx$ normal for many distributions.

For many distributions $\min(Y_1, \ldots, Y_n) \approx$ Weibull.

Suppose a product is made of m components with iid failure times T_i. Suppose the product fails as soon as one of the components fails, eg the weakest link in a chain $\infty\infty\infty\ldots\infty$. Then often failure time $T = \min(T_1, \ldots, T_m) \approx$ Weibull.

7) Notation $\{t : f(t) > 0\}$ = support of f.

If the support is $t > 0 = (0, \infty)$, often formulas will omit $t > 0$. For ex $h(t) = \lambda$, $H(t) = \lambda t$, $S(t) = F(t) = 1 - e^{-\lambda t}$ and $f(t) = \lambda e^{-\lambda t}$ for $T \sim EXP(\lambda)$.

5) The Gompertz distribution has $h(t) = \lambda e^{\delta t}$.

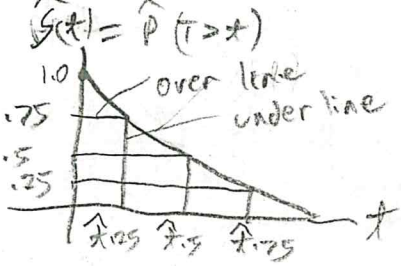
6) The loglogistic, lognormal gamma and inverse Gaussian distributions are sometimes used.

Ch2 Some nonparametric procedures see Olive@siu.edu

1) Know for exam! Let $\hat{S}(t)$ be the estimated survival function. Let $t(p)$ be the pth percentile: $P(T \leq t(p)) = p = F(t(p))$. So $1 - p = S(t(p)) = P(T > t(p))$. Given a graph of $S(t)$ or $\hat{S}(t)$, to estimate $\hat{t}(p)$, when $100(p)\%$ have died, use "over" and "down" lines.

a) Find $1-p$ on the vertical axis. Draw a horizontal "over" line until it intersects $\hat{S}(t)$.

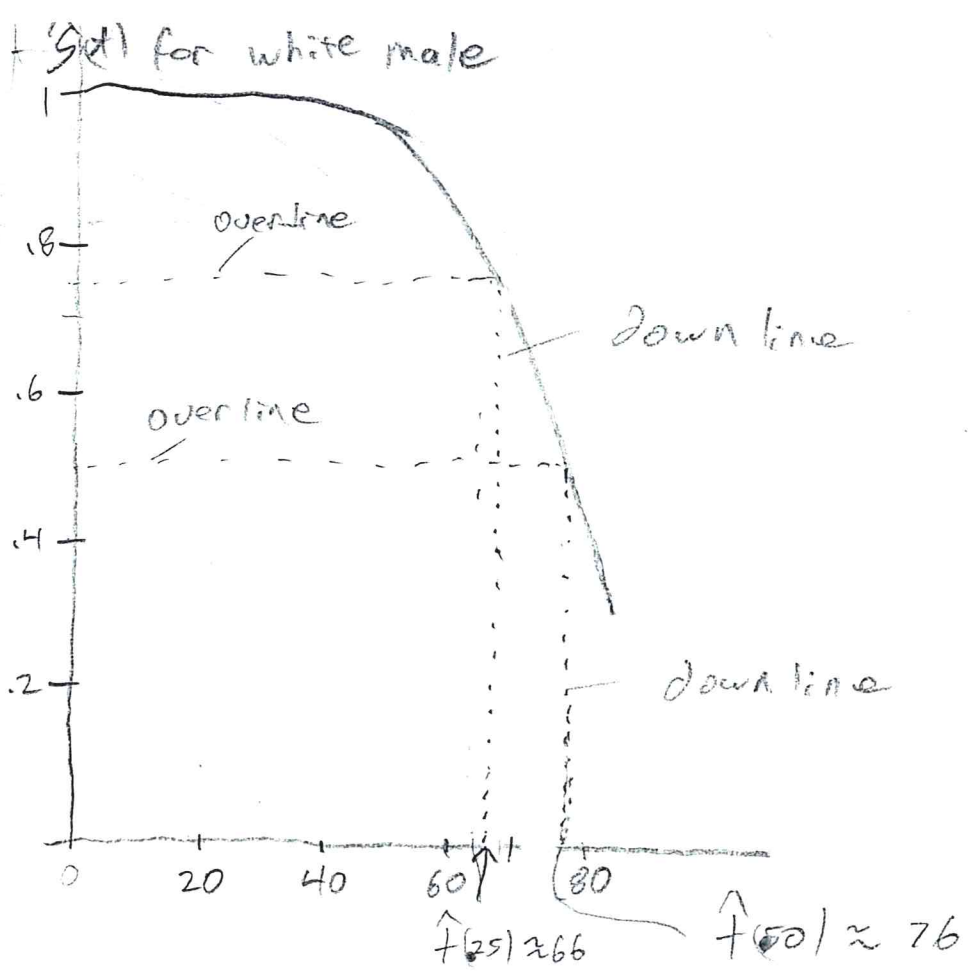
b) Draw vertical "down" line until it intersects the horizontal axis at $\hat{t}(p)$.



Usually want $\hat{t}(.5)$ = 50% dead

but sometimes want $\hat{t}(.25)$ = 25% dead or $\hat{t}(.75)$ = 75% dead

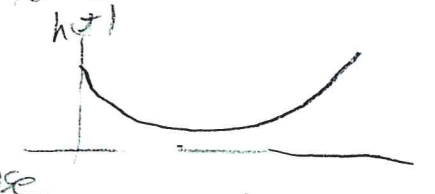
ex) handout gives $\hat{S}(t)$ for $t = 0, 1, \ldots, 85$ years for black and white males and females in US 1989.



Returning to handout, when have about 1% of people died? A at age 1. According to Elandt Johnson and Johnson, p86, $\frac{3}{4}$ of babies that die in the 1st year die in the 1st week. Birth defects last until about 10. For white females when do the next 1% die? A $\approx$ 33 (44% alive 66% dead)

If a white male lives to 78, he has about a 50% chance of living to 85. A white female who lives to 68 has about a 50% chance of living to 85.

$h(t)$ is bath tub shaped.



For young adults car accidents homicide suicide heart disease and cancer are the big killers. The rates are much higher for males.

Some types of microchips have lots of initial defects. use "burn in" run current through chips for 1 week, say,

Lab 1 www.[illegible]/olive/students.htm 1-5

HW 2.1 SAS Show word, but print from SAS

plots(s,h) to plots(h) will print hazard function

HW 3.02 SAS KM

www.math.siu.edu/olive/infboot.htm → sipack.txt
copy accisim which is classical US ≈ plus 4 for

R

accisim(n=10, nruns=100, p=.5)

want cov > .93 and len ≈ 1.96 for large n
small len is good

n	p	
10	.5	
100	.5	
10		.0001
100		
1000		
100	.01	
1000	.01	
100	.99	
1000	.99	
1000	.5	
10000	.5	

do up arrows several times

zy
zz
z+
zdelq
out
st

where y_2 is held fixed (get the region of integration, draw a line parallel to the y_1 axis and use the functions $y_1 = \psi_L(y_2)$ and $y_1 = \psi_R(y_2)$ as the lower and upper limits of integration). The **conditional probability density function of Y_1 given $Y_2 = y_2$** is

$$f_{Y_1|Y_2=y_2}(y_1|y_2) = \frac{f(y_1, y_2)}{f_{Y_2}(y_2)}$$

provided $f_{Y_2}(y_2) > 0$. The **conditional probability density function of Y_1 given $Y_2 = y_2$** is

$$f_{Y_2|Y_1=y_1}(y_2|y_1) = \frac{f(y_1, y_2)}{f_{Y_1}(y_1)}$$

provided $f_{Y_1}(y_1) > 0$.

The **support** of a continuous RV is $\{y : f(y) > 0\}$. The support of jointly continuous (Y_1, Y_2) is $\{(y_1, y_2) : f(y_1, y_2) > 0\}$.

The **support** of a discrete RV is $\{y : p(y) > 0\}$. The support of jointly discrete (Y_1, Y_2) is $\{(y_1, y_2) : p(y_1, y_2) > 0\}$.

The *support* of the conditional probability function or pdf can depend on the 2nd variable. For example, the support of $f_{Y_1|Y_2=y_2}(y_1|y_2)$ could have the form $0 \leq y_1 \leq y_2$.

Material after exam 2.

Random variables Y_1 and Y_2 are **independent** if any one of the following conditions holds.
i) $F(y_1, y_2) = F_{Y_1}(y_1)F_{Y_2}(y_2) \quad \forall y_1, y_2$.
ii) $p(y_1, y_2) = p_{Y_1}(y_1)p_{Y_2}(y_2) \quad \forall y_1, y_2$.
iii) $f(y_1, y_2) = f_{Y_1}(y_1)f_{Y_2}(y_2) \quad \forall y_1, y_2$.
Otherwise, Y_1 and Y_2 are *dependent.*

If $Y_1, Y_2, ..., Y_n$ are independent if $\forall y_1, y_2, ..., y_n$:
i) $F(y_1, y_2, ..., y_n) = F_{Y_1}(y_1)F_{Y_2}(y_2) \cdots F_{Y_n}(y_n)$
ii) $p(y_1, y_2, ..., y_n) = p_{Y_1}(y_1)p_{Y_2}(y_2) \cdots p_{Y_n}(y_n)$ or
iii) $f(y_1, y_2, ..., y_n) = f_{Y_1}(y_1)f_{Y_2}(y_2) \cdots f_{Y_n}(y_n)$.

Two RV's Y_1 and Y_2 are dependent if their support is not rectangular. If the support is rectangular, another test must be used to determine whether Y_1 and Y_2 are independent or dependent.

If continuous Y_1 and Y_2 have rectangular support, then Y_1 and Y_2 are independent iff $f(y_1, y_2) = g(y_1)h(y_2)$ for (y_1, y_2) in the support where g is a nonnegative function of y_1 alone and h is a nonnegative function of y_2 alone.

To check whether discrete Y_1 and Y_2 (with rectangular support) are independent given a 2 by 2 table, find the row and column sums and check whether $p(y_1, y_2) \neq p_{Y_1}(y_1)p_{Y_2}(y_2)$ for **some entry** (y_1, y_2). Then Y_1 and Y_2 are dependent. If $p(y_1, y_2) = p_{Y_1}(y_1)p_{Y_2}(y_2)$ for *all table entries, then Y_1 and Y_2 are independent.*

3) **Common Problem.** Determine whether Y_1 and Y_2 are independent or dependent. Q6 1a, 2a, HW10 5.39, 5.42, 5.44, 5.45, 5.53

2) * The empirical survival function

$$\hat{S}_E(t) = \frac{\text{number of individuals with survival times} \geq t}{\text{number of individuals}}$$

(for i.i.d data)

and is useful if there is no censoring.

warning: Collett uses $\geq$ not $>$.

3) Let the indicator random variable

$$W_i = I_A(T_i) = \begin{cases} 1 & T_i \in A \\ 0 & T_i \notin A \end{cases}$$

Sometimes write $I_{(t,\infty)}(T_i)$ as $I(T_i > t)$.

4) $$\hat{S}_E(t) = \frac{1}{n}\sum_{i=1}^{n} I_{(t,\infty)}(T_i) = \frac{1}{n}\sum_{i=1}^{n} I(T_i > t) = \hat{P}_t$$

where $\hat{P}_t$ is the sample proportion of lifetimes $> t$.

5) A <u>distribution free</u> or <u>nonparametric</u> estimate makes few assumptions on the T_i, eg $T_1, \ldots, T_n$ are iid with $T_i > 0$.

This chapter covers nonparametric estimators.

A parametric model specifies the distribution up to a few unknown parameters, eg $T_1, \ldots, T_n$ are iid $EXP(\lambda)$.

6) Assume $T_1, \ldots, T_n$ are iid with $T_i > 0$. Fix t.

Then $I(T_i > t)$ are iid binomial $[1, p = P(T > t)]$

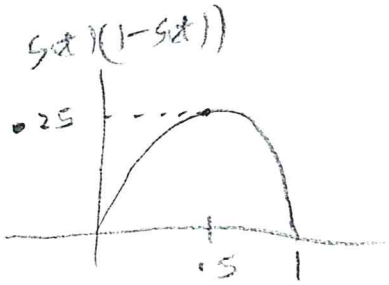
So $n\hat{S}_E(t) \sim$ binomial $(n, p = P(T > t))$.

Hence $E(n\hat{S}_E(t)) = nP(T > t)$ and $V(n\hat{S}_E(t)) = nS(t)F(t)$.

Recall that $E(cW) = cE(W)$ and $V(cW) = c^2 V(W)$.

Hence $E(\hat{S}_E(t)) = P(T > t) = S(t)$ and

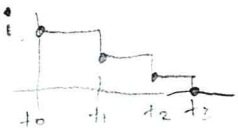
$$V(\hat{S}_E(t)) = \frac{S(t)F(t)}{n} = \frac{S(t)(1-S(t))}{n} \le \frac{.25}{n}$$



$$SD(\hat{S}_E(t)) = \sqrt{V(\hat{S}_E(t))} \le \frac{.5}{\sqrt{n}}$$

so need $n \approx 100$ for $SD < .05$.

7) know for exam 1 Let $t_{(1)} \le t_{(2)} \le \cdots \le t_{(n)}$ be the observed lifetimes. Let $0 < t_1 < t_2 < \cdots < t_m$ be the distinct observed lifetimes. Let d_i = # of deaths at time t_i. If $m = n$ and $d_i \equiv 1$ for $i = 1, \ldots, n$, then there are no ties. If $m < n$ and some $d_i \ge 2$, then there are ties. $\hat{S}_E(t)$ is a step function with jumps at t_i, $n = \sum_{i=1}^{m} d_i$



$\hat{S}_E(t_0) = \hat{S}_E(0) = 1$, $\hat{S}_E(t) = \hat{S}_E(t_{i-1})$ for $t_{i-1} \le t < t_i$

t_i	d_i	$\hat{S}_E(t_i) = \frac{a_i}{n} = \hat{S}_E(t_{i-1}) - \frac{d_i}{n}$ where $a_i = \#\, T_i > t_i = \sum_{k=1}^{n} I(T_i > t_i)$, $i = 1, \ldots, m$, $a_0 = n$
$t_0 = 0$		$\hat{S}_E(t_0) = 1 = \frac{n}{n} = \frac{a_0}{n}$
t_1	d_1	$\frac{n-d_1}{n} = \hat{S}_E(t_0) - \frac{d_1}{n} = \hat{S}_E(t_1) = \frac{a_1}{n}$
t_2	d_2	$\frac{n-d_1-d_2}{n} = \hat{S}_E(t_1) - \frac{d_2}{n} = \hat{S}_E(t_2) = \frac{a_1-d_2}{n} = \frac{a_2}{n}$
$\vdots$		
t_j	d_j	$\frac{n-d_1-d_2-\cdots-d_j}{n} = \hat{S}_E(t_{j-1}) - \frac{d_j}{n} = \hat{S}_E(t_j) = \frac{a_{j-1}-d_j}{n} = \frac{a_j}{n}$
$\vdots$		
t_{m-1}	d_{m-1}	$\frac{n-d_1-\cdots-d_{m-1}}{n} = \hat{S}_E(t_{m-2}) - \frac{d_{m-1}}{n} = \hat{S}_E(t_{m-1}) = \frac{a_{m-2}-d_{m-1}}{n} = \frac{a_{m-1}}{n}$
t_m	d_m	$\frac{n-\sum d_i}{n} = \frac{n-n}{n} = 0 = \hat{S}_E(t_m) = \frac{a_{m-1}-d_m}{n} = \frac{d_m-d_m}{n} = \frac{a_m}{n}$

$a_{m-1} = d_m$ and $a_m = 0$

know for ≠1 given d_i and t_i $i=1,\dots,m$ $t_1 < t_2 < \dots < t_m$

or given $t_{(1)} \le t_{(2)} \le \dots \le t_{(n)}$, be able to compute and plot $\hat{S}_E(t)$ for $t > 0$.

$n = 21$ Steroid induced remission times for leukemia patients

Placebo group time in weeks

ex) needed work

$a_i = \#>t_i$	$t_{(j)}$	t_i	d_i	$\hat{S}_E(t_i) = \hat{S}_E(t_{i-1}) - \frac{d_i}{n}$
21	(0	$t_0 = 0$		$\hat{S}_E(0) = 1 = \frac{21}{21}$
	1			
19	1	$t_1 = 1$	2	$\hat{S}_E(1) = \frac{21-2}{21} = \frac{19}{21}$
	2			
17	2	$t_2 = 2$	2	$\hat{S}_E(2) = \frac{19-2}{21} = \frac{17}{21}$
16	3	$t_3 = 3$	1	$\hat{S}_E(3) = \frac{17-1}{21} = \frac{16}{21}$
	4			
14	4	$t_4 = 4$	2	$\hat{S}_E(4) = \frac{16-2}{21} = \frac{14}{21}$
	5			
12	5	$t_5 = 5$	2	$\hat{S}_E(5) = \frac{14-2}{21} = \frac{12}{21}$
	8			
	8			
	8			
8	8	$t_6 = 8$	4	$\hat{S}_E(8) = \frac{12-4}{21} = \frac{8}{21}$
	11			
6	11	$t_7 = 11$	2	$\hat{S}_E(11) = \frac{8-2}{21} = \frac{6}{21}$
	12			
4	12	$t_8 = 12$	2	$\hat{S}_E(12) = \frac{6-2}{21} = \frac{4}{21}$
3	15	$t_9 = 15$	1	$\hat{S}_E(15) = \frac{4-1}{21} = \frac{3}{21}$
2	17	$t_{10} = 17$	1	$\hat{S}_E(17) = \frac{3-1}{21} = \frac{2}{21}$
1	22	$t_{11} = 22$	1	$\hat{S}_E(22) = \frac{2-1}{21} = \frac{1}{21}$
0	23	$t_{12} = 23$	1	$\hat{S}_E(23) = \frac{1-1}{21} = \frac{0}{21}$ ← good check

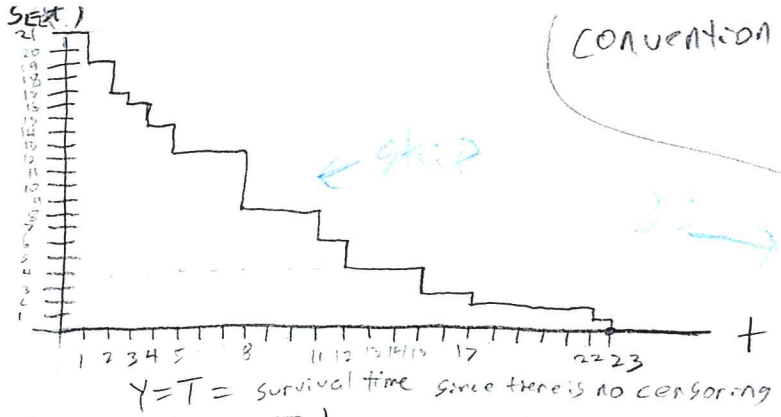
write small!

Smith p68

good check: LHS $= \frac{a_i}{n}$ (RHS $= \frac{a_i}{n}$)

not needed if t_i and d_i are given

$t_{(1)}, \dots, t_{(n)}$ ← notation is confusing, depends on whether n times or m times are given

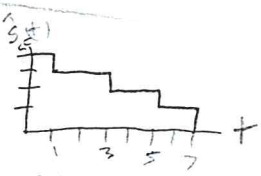


convention: draw in vertical lines

ex) t_i: 1, 3, 5, 7

d_i: 1, 1, 1, 1

$\frac{3}{4}$, $\frac{3}{4}\cdot\frac{1}{2}$, $a_1 = 3$, $a_2 = 2$, $a_3 = 1$



9) know for E1 Let $t_0 = 0 < t_1 < \cdots < t_m$ be the distinct death times. Let $t_1 \le t_c < t_m$. Then the classical large sample 95% CI for $S_Y(t_c)$ is

$$\hat{S}_E(t_c) \pm 1.96\sqrt{\frac{\hat{S}_E(t_c)\left(1-\hat{S}_E(t_c)\right)}{n}}$$

$= \hat{S}_E(t_c) \pm 1.96\, SE\left(\hat{S}_E(t_c)\right)$ $= (L, U)$ use $(\max(0,L), \min(1,U))$.

Note that the CI has the form $\hat{p}_{t_c} \pm 1.96\sqrt{\frac{\hat{p}_{t_c}(1-\hat{p}_{t_c})}{n}}$.

10) know for E1. Another approximate 95% CI for $S_Y(t_c)$ is $\tilde{p}_{t_c} \pm 1.96\sqrt{\frac{\tilde{p}_{t_c}(1-\tilde{p}_{t_c})}{n+4}}$ $= (L, U)$. use $= (\max(0,L), \min(1,U))$; where $t_c > 0$

and $\tilde{p}_{t_c} = \frac{n\,\hat{S}_E(t_c) + 2}{n+4}$. This is called a plus four 95% CI for $S_Y(t_c)$. (4 imaginary T_i^* are added to the sample. Two of the T_i^* are $> t_m > t_c$ and two are $< t_1 < t_c$.)

ex) Let $n = 21$ $\hat{S}_E(12) = \frac{4}{21}$.

a) Find the 95% classical CI for $S(12)$.

b) Find the 95% plus four CI for $S(12)$.

soln a) $\hat{S}_E(12) \pm 1.96 \sqrt{\frac{\hat{S}_E(12)(1-\hat{S}_E(12))}{n}}$

$= \frac{4}{21} \pm 1.96 \sqrt{\frac{\frac{4}{21}(1-\frac{4}{21})}{21}} = \frac{4}{21} \pm .16795$

$= (.0225, .3584)$

b) $\tilde{p}_{12} = \frac{n \hat{S}_E(12) + 2}{n+4} = \frac{21 \frac{4}{21} + 2}{21+4} = \frac{6}{25}$

So the 95% CI is $\tilde{p}_{12} \pm 1.96 \sqrt{\frac{\tilde{p}_{12}(1-\tilde{p}_{12})}{n+4}}$

$= \frac{6}{25} \pm 1.96 \sqrt{\frac{\frac{6}{25}(1-\frac{6}{25})}{25}} = \frac{6}{25} \pm .16742$

$= (.0726, .4074)$

So .07 to .41 are reasonable values for $S(12)$.
The CI is not very precise (short) because n is small

11) * Suppose $T \geq 0$ and $[0, \infty) = \underbrace{[t_0, t_1)}_{I_1} \cup \underbrace{[t_1, t_2)}_{I_2} \cup \cdots \cup \underbrace{[t_{m-1}, t_m)}_{I_m}$, $t_0 = 0$, $t_m = \infty$

Let Y_i = time to event and $T_i = \min(Y_i, Z_i)$.
Want to estimate $S_Y(t) = P(Y > t)$, but there is censoring.

Let $p_j = P$ (Surviving through I_j | Alive at the start of I_j)

$= P(Y > t_j \mid Y > t_{j-1}) = \dfrac{P(Y > t_j,\ Y > t_{j-1})}{P(Y > t_{j-1})} = \dfrac{S(t_j)}{S(t_{j-1})}$.

So $p_1 = \dfrac{S(t_1)}{S(t_0)} = S(t_1) \equiv S_Y(t_1)$ since $S(0) = S(t_0) = 1$.

$\therefore S_Y(t_K) \equiv S(t_K) = S(t_1)\ \dfrac{S(t_2)}{S(t_1)}\ \dfrac{S(t_3)}{S(t_2)} \cdots \dfrac{S(t_{K-1})}{S(t_{K-2})}\ \dfrac{S(t_K)}{S(t_{K-1})}$ (telescoping product)

$= p_1 p_2 \cdots p_K = \prod_{j=1}^{K} p_j$.

Let $\hat{p}_j = 1 - \dfrac{\text{number dying in } I_j}{\text{number with potential to die in } I_j}$

(if 5/50 die then 45/50 live)

A = surviving in I_j | alive; dying in I_j | alive

$\hat{P}(A) = 1 - \hat{P}(\bar{A})$ (complement rule)

12) p17 Suppose that the actual death and censoring times are unknown, but we know the number of deaths and censoring times in each interval I_j, $j = 1, \ldots, m$.

Let d_j = # deaths in I_j

c_j = # of censored survival times in I_j.

n_j = # at risk in I_j = # who were alive and uncensored at time t_{j-1}, eg at the start of the interval. Note that $n_1 = n$ and $n_j = n_{j-1} - d_{j-1} - c_{j-1}$ for $j > 1$. This shows how those at risk in the $(j-1)$th interval propagate into the jth interval.

If everyone is censored at the start of I_j, then the number at risk is $n_j - c_j$. If everyone is censored at the end of the interval then the number at risk is n_j.

Let $n_j' = n_j - \dfrac{c_j}{2} = \dfrac{n_j + n_j - c_j}{2}$ = average # at risk in interval I_j (eg if censoring is uniform over the interval). Then $\tilde{p}_j = 1 - \dfrac{d_j}{n_j'} = \dfrac{n_j' - d_j}{n_j'}$.

13) Know the lifetable of actuarial method for estimating $S_Y(t)$ is $\hat{S}_L(0) = 1$,

$$\hat{S}_L(t_k) = \prod_{j=1}^{k} \left(\frac{n_j' - d_j}{n_j'} \right) = \prod_{j=1}^{k} \tilde{p}_j$$

for $k = 1, \ldots, m-1$. If $t_m = \infty$ $\hat{S}_L(t)$ is undefined for $t > t_{m-1}$. If $t_m \neq \infty$ take $\hat{S}_L(t) = 0$, $t \geq t_m$.

To graph $\hat{S}_L(t)$ use linear interpolation (connect the dots).

If $n_j' = 0$, take $\tilde{p}_j = \frac{n_j' - d_j}{n_j'} = 0$.

Note that $\hat{S}_L(t_k) = \hat{S}_L(t_{k-1}) \frac{n_k' - d_k}{n_k'}$ for $k = 1, \ldots, m$

see 9) in EI review

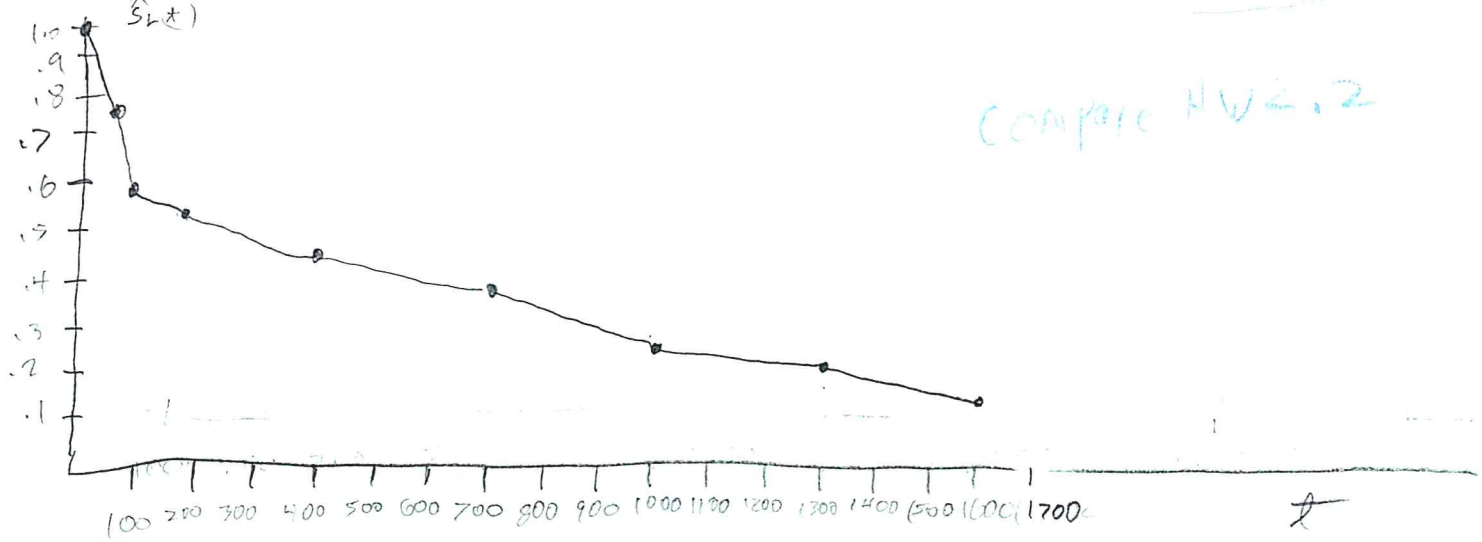
	interval I_j	d_j	c_j	n_j	n_j'	$\frac{n_j' - d_j}{n_j'}$	$\hat{S}_L(t)$
1	$[t_0, t_1)$ ($t_0 = 0$)	d_1	c_1	n_1	$n_1 - \frac{c_1}{2}$	$\frac{n_1' - d_1}{n_1'}$	$\hat{S}_L(0) = \hat{S}_L(t_0) = 1$
2	$[t_1, t_2)$	d_2	c_2	n_2	$n_2 - \frac{c_2}{2}$	$\frac{n_2' - d_2}{n_2'}$	$\hat{S}_L(t_1) = \hat{S}_L(t_0) \frac{n_1' - d_1}{n_1'}$
3	(t_2, t_3)	d_3	c_3	n_3	$n_3 - \frac{c_3}{2}$	$\frac{n_3' - d_3}{n_3'}$	$\hat{S}_L(t_2) = \hat{S}_L(t_1) \frac{n_2' - d_2}{n_2'}$
k	$[t_{k-1}, t_k)$	d_k	c_k	n_k	$n_k - \frac{c_k}{2}$	$\frac{n_k' - d_k}{n_k'}$	$\hat{S}_L(t_{k-1}) = \hat{S}_L(t_{k-2}) \frac{n_{k-1}' - d_{k-1}}{n_{k-1}'}$
$m-1$	$[t_{m-2}, t_{m-1})$	d_{m-1}	c_{m-1}	n_{m-1}	$n_{m-1} - \frac{c_{m-1}}{2}$	$\frac{n_{m-1}' - d_{m-1}}{n_{m-1}'}$	$\hat{S}_L(t_{m-2}) = \hat{S}_L(t_{m-3}) \frac{n_{m-2}' - d_{m-2}}{n_{m-2}'}$
m	$[t_{m-1}, t_m)$ ($t_m = \infty$)	d_m	c_m	n_m			$\hat{S}_L(t_{m-1}) = \hat{S}_L(t_{m-2}) \frac{n_{m-1}' - d_{m-1}}{n_{m-1}'}$

convention: If $t_0 \neq 0$, find $\hat{S}(t_1)$ and graph from $(0, \hat{S}(0) = 1)$ to $(t_1, \hat{S}(t_1))$ instead of using a

if $t_m \neq \infty$

ex 68 heart transplant patients (Allison)

I_j	t_j	d_j	c_j	n_j	$n_j - \frac{c_j}{2}$	$\frac{n_j - d_j}{n_j}$	$\hat{S}_L(t_j) = \hat{S}(t_{j-1})\frac{n_j - d_j}{n_j}$
$[0,50)$	0	16	3	68	66.5	.7594	$\hat{S}(0) = 1$
$[50,100)$	50	11	0	49	49	.7755	$\hat{S}(50) = 1(.7594) = .7594$
$[100,200)$	100	4	2	38	37	.8919	$\hat{S}(100) = .7594(.7755) = .588$
$[200,400)$	200	5	4	32	30	.8333	$\hat{S}(200) = .5889(.8919) = .525$
$[400,700)$	400	2	6	23	20	.90	$\hat{S}(400) = .5252(.8333) = .4376$
$[700,1000)$	700	4	3	15	13.5	.7037	$\hat{S}(700) = .4376(.90) = .3938$
$[1000,1300)$	1000	1	2	8	7	.8571	$\hat{S}(1000) = .3938(.7037) = .2771$
$[1300,1600)$	1300	1	3	5	3.5	.7143	$\hat{S}(1300) = .2771(.8571) = .2375$
$[1600,\infty)$	1600	0	1	1	.5		$\hat{S}(1600) = .2375(.7134) = .1696$



compare HW 2.2

14) Greenwood's formula

$$SE\left(\hat{S}_L(t_j)\right) = \hat{S}_L(t_j)\sqrt{\sum_{i=1}^{j}\frac{1-\tilde{p}_i}{\tilde{p}_i n_i'}} \approx \hat{S}_L(t_j)\sqrt{\sum_{i=1}^{j}\frac{d_i}{n_i'(n_i'-d_i)}}$$

where $\tilde{p}_i \approx \frac{n_i' - d_i}{n_i'}$ and $j = 1, \ldots, m-1$. $\left(1-\tilde{p}_i = \frac{d_i}{n_i'} \text{ and } \tilde{p}_i n_i' = n_i' - d_i\right)$

ex) $SE\left(\hat{S}_L(100)\right) = SE\left(\hat{S}_L(t_2)\right) = .5889\sqrt{\frac{1-.7594}{.7594(66.5)} + \frac{1-.7755}{.7755(49)}}$

$= .0608$ Usually obtained from output.

15) * using output

Interval	survival	survival SE
0 50	1.00	0
50 100	.7594	.0524
100 200	.5889	.0608

$\hat{S}(0) = 1$

$\hat{S}(50) = .7594$ $\quad SE\ \hat{S}(50) = .0524$

or output

Interval	survival	survival SE
0 50	.7594	.0524
50 100	.5889	.0608
100 200	.5253	.0602

You know $\hat{S}(0) = 1$, so $\hat{S}_L(50) = .7594$, $SE(\hat{S}_L(50)) = .0524$

Output is for left or right endpoint, to tell which endpoint is used, look at the interval $(0, t_i)$

16) p26 know for exam! Be able to get a 95% CI for $S_Y(t_i)$ from output

$= \hat{S}_L(t_i) \pm 1.96\ SE\ \hat{S}_L(t_i) = (L, U)$

then use $(\max(0, L), \min(U, 1))$

ex) A 95% CI for $S_Y(50)$ is

$.7594 \pm 1.96(.0524) = (.6567, .8621)$

17) *or output is like (survival distribution function, upper confidence limit)

time	survival	SDF_LCL	SDF_UCL
0	1.0	1.0	1.0
50	.7594	.65666	.86213

Then the 95% CI for S(50) is (.65666, .86213)

18) $Y_i^* = T_i = \min(Y_i, Z_i)$ Y_i and Z_i are ind

$\delta_i = \begin{cases} 0 & \text{censored } T_i = Z_i \\ 1 & \text{uncensored } T_i = Y_i \end{cases} = I(Y_i \le Z_i)$

$t_{(1)} \le t_{(2)} \le \cdots \le t_{(n)}$ ordered times

γ_1 γ_2 γ_n

$\gamma_i = \begin{cases} 0 & \text{if } t_{(i)} \text{ is censored} \\ 1 & \text{if } t_{(i)} \text{ is uncensored} \end{cases}$

$t_i, i=1,\dots,m$ = ordered distinct uncensored times, $i=1,\dots,m$

ex) $n=6$ $Y_i \sim EXP(1)$ $E(Y_i)=1$

$Z_i \sim EXP(.1)$ $E(Z_i)=10$

Y_i	.2887	.1796	1.1301	1.4165	.2720	.6667	not observed except in computer simulation
Z_i	.8967	1.6158	10.5266	1.052	2.2329	4.2917	
$T_i = Y_i^*$	.2887	.1796	1.1301	1.052	.2720	.6667	
δ_i	1	1	1	0	1	1	
$t_{(j)}$	.1796	.2720	.2887	.6667	1.0522	1.1301	
γ_j	1	1	1	1	0	1	
t_i	.1796	.2720	.2887	.6667		1.1301	$m=5$

19) Consider intervals $(0, t_{(1)}]$ $(t_{(1)}, t_{(2)}]$ $\cdots$ $(t_{(n-1)}, t_{(n)}]$

$I_{(1)}$ $I_{(2)}$ $I_{(n)}$

Let $n_{(j)}$ = # at risk in $I_{(j)}$ $d_{(j)}$ = # deaths and

$r(j)$ – 1 (survive through $t_{(j)}$) divide at risk at $t_{(j)}$

Then $p^*_{(j)} = 1 - \frac{\#\text{ dying in } I_j}{\#\text{ with potential to die in } I_j} = \begin{cases} 1 - \frac{d_{(j)}}{n_{(j)}}, & \gamma_j = \\ 1 - \frac{0}{n_{(j)}} = 1, & \gamma_j = \end{cases}$

and $S^*(t_{(j)}) = \prod_{i=1}^{j} p^*_{(j)}$.

Want to use t_i instead of $t_{(j)}$ since $p^*_{(j)} = 1$ if $\gamma_j = 0$.

20) Let $t_1, \ldots, t_m$ be the ordered times where an event (death) occurred.

Let n_i = # at risk at t_i = # alive and uncensored just before t_i
$n_i = \sum_{j=1}^{n} I(t_j \geq t_i)$,
Let d_i = # of events (deaths) at t_i. Then $d_i \geq 1$

and if $d_i = 1$ for $i = 1, \ldots, m$ there are no ties

If $d_i > 1$ for some i, then there are ties.

Note that individuals who die or are censored at time t_i are "at risk at t_i". So $n_1 = n$ if $t_1 = t_{(1)}$

In interval $(t_{i-\delta}, t_i]$, conditional prob of dying = $\frac{d_i}{n_i}$ and conditional prob of surviving = $1 - \frac{d_i}{n_i}$.

21) know for Final p20 Let $t_0 = 0$. Then the Kaplan Meier

p28-32 estimator or product limit estimator of

$S_Y(t_i) = P(Y > t_i)$ is $\hat{S}_{KM}(0) = 1$ and

$$\hat{S}_{KM}(t_i) = \prod_{k=1}^{i} \left(1 - \frac{d_k}{n_k}\right) = \hat{S}_{KM}(t_{i-1})\left(1 - \frac{d_i}{n_i}\right).$$

$\hat{S}_{KM}(t)$ is a step function: $\hat{S}_{KM}(t) = \hat{S}_{KM}(t_{i-1})$ for $t_{i-1} \leq t <$

If $t_{(n)}$ is uncensored $\hat{S}_{KM}(t) = 0$ for $t > t_{(n)}$,

If $t_{(n)}$ is censored $\hat{S}_{KM}(t) = \hat{S}_{KM}(t_m)$ for $t_m \leq t < t_{(n)}$, but $\hat{S}_{KM}(t)$ is undefined for $t > t_{(n)}$

22) know for exam 1 given $(t_{(j)}, \gamma_j)$ be able

p34-35

to compute $\hat{S}_K(t)$ from a table. Convention

if a death and censored obs are tied, put the

censored obs after the death. eg 10, 10+

10	1
10+	0

Let $n_0 = n$:

$n_i = n_{i-1} - f_{i-1} = \#\, t_{(j)} \geq t_i$, (eg $n_1 = n$ if no cases were censored before t_1)

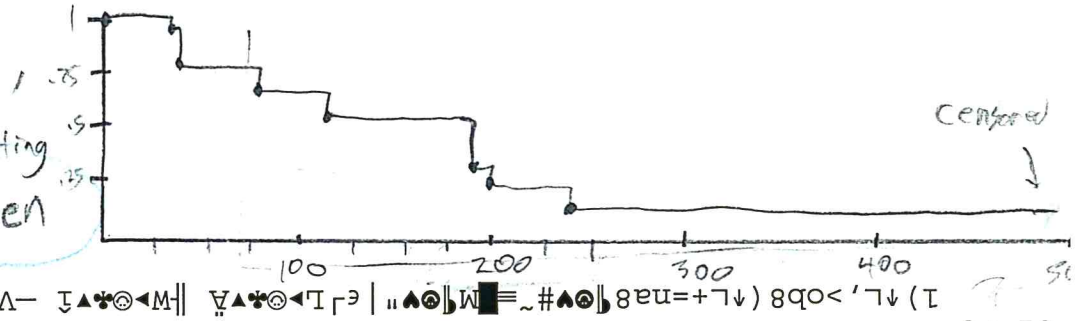
f_{i-1} = # deaths and number censored in $[t_{i-1}, t_i)$

Survival time in days until repair of n=13 street lights (Smith p113) modified

ex) 36, 38, 38, 38+, 78, 112, 112, 114+, 162+, 189, 198, 237, 489+

f_i	$t_{(j)}$	γ_j	t_i	n_i	d_i	$\hat{S}_K(t_i) = \hat{S}_K(t_{i-1})\left(1 - \frac{d_i}{n_i}\right)$
			$t_0=0$			$\hat{S}(0) = \hat{S}(t_0) = 1$
f_1 1	36	1	t_1	36	13	1	$\hat{S}(36) = 1\left(1 - \frac{1}{13}\right) = \frac{12}{13} \approx .9231$
f_2 3	38	1	t_2	38	12	2	$\hat{S}(38) = (.9231)\left(1 - \frac{2}{12}\right) \approx .7692$
	38	1					
	38	0					
f_3 1	78	1	t_3	78	9	1	$\hat{S}(78) = (.7692)\left(1 - \frac{1}{9}\right) = .6837$
f_4 4	112	1	t_4	112	8	2	$\hat{S}(112) = .6837\left(1 - \frac{2}{8}\right) = .5128$
	112	1					
	114	0					
	162	0					
f_5 1	189	1	t_5	189	4	1	$\hat{S}(189) = .5128\left(1 - \frac{1}{4}\right) = .3846$
f_6 1	198	1	t_6	198	3	1	$\hat{S}(198) = .3846\left(1 - \frac{1}{3}\right) = .2564$
f_7 1	237	1	t_7	237	2	1	$\hat{S}(237) = .2564\left(1 - \frac{1}{2}\right) = .1282$
	489	0					

not needed except for plotting if t_i, n_i and d_i are given



if 489 was uncensored $\hat{S}_K(t) = 0$ for $t \geq 489$

since 489 is censored $\hat{S}_K(t)$ is undefined for $t \geq 489$

23) p42 Greenwood's formula.

$$SE\left(\hat{S}_K(t_j)\right) = \hat{S}_K(t_j)\sqrt{\sum_{i=1}^{j} \frac{d_i}{n_i(n_i - d_i)}}$$

is usually obtained from output.

24) know for E1 Be able to find a 95% CI for $S_Y(t_i)$ from output. (See points 15) 16) and 17)). $\hat{S}_K(t_i) \pm 1.96\, SE\left(\hat{S}_K(t_i)\right)$

p26 $=(L,U)$ then use $(\max(0,L), \min(1,U))$.

ex) R output

t_i time	n_i n.risk	d_i n.event	$\hat{S}(t_i)$ survival	$SE(\hat{S}(t_i))$ std err	lower 95%CI	upper 95%CI
36	13	1	.923	.0739	.7782	1.00

95%CI for $S_Y(36)$ is is $(.7782, 1)$

or $\hat{S}_K(36) \pm 1.96\, SE\left(\hat{S}_K(36)\right) = .923 \pm 1.96(.0739)$

$= .923 \pm .1448 = (.7782, 1.0678)$

but $S_Y(36) \in (0,1)$ so use $(.7782, 1)$.

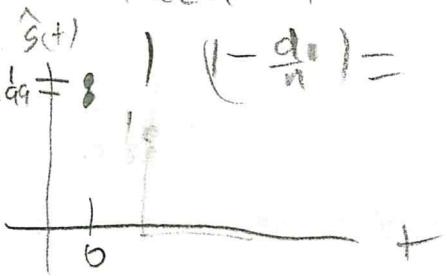
(Similarly round negative lower limit up to 0.)

Skip to 34) for HW3

25) It is possible, due to rounding, that $t_1 = 0$. For ex $n = n_1 = 100$, $t_1 = 0$, $d_1 = 1$.

Then take $\hat{S}_K(0) = 1$, $\hat{S}_K(t_1) = \hat{S}_K(\delta) = 1\left(1 - \frac{d_1}{n}\right) = 1 - \frac{1}{100} = .99$. Here $\delta > 0$ is tiny.

Graph $\hat{S}(0)$ and $\hat{S}(\delta)$ at $t = 0$.



26) * The Nelson Aalen estimator of $S_Y(t)$ is

$$\hat{S}_N(t_i) = \prod_{k=1}^{i} \exp\left(-\frac{d_k}{n_k}\right) = \exp\left(-\sum_{k=1}^{i} \frac{d_k}{n_k}\right)$$

$$= \hat{S}_N(t_{i-1}) \exp\left(-\frac{d_i}{n_i}\right) \quad \text{where } \hat{S}(0) = 1$$

and t_i, d_i and n_i are the same as for the Kaplan Meier estimator.

27) * p75 The Kaplan Meier estimator of $H_Y(t)$ is

$\hat{H}_K(0) = 0$, $\hat{H}_K(t_i) = -\log \hat{S}_K(t_i) = -\sum_{k=1}^{i} \log\left(1 - \frac{d_k}{n_k}\right)$

$= \hat{H}_K(t_{i-1}) - \log\left(1 - \frac{d_i}{n_i}\right)$.

28) * p75 The Nelson Aalen estimator of $H_Y(t)$ is

$\hat{H}_N(0) = 0$, $\hat{H}_N(t_i) = \sum_{k=1}^{i} \frac{d_k}{n_k} = \hat{H}_N(t_{i-1}) + \frac{d_i}{n_i}$.

Note that $\hat{S}_N(t_i) = \exp(-\hat{H}_N(t_i))$ so $\hat{H}_N(t) = -\log(\hat{S}_N(t_i))$

29) know for E1 A 95% CI for $H_Y(t_i)$ is

$\hat{H}(t_i) \pm 1.96\, SE(\hat{H}(t_i))$. $= (L, U)$. use $(\max(0, L), U)$.

30) $SE(\hat{H}_N(t_i)) = \sqrt{\sum_{k=1}^{i} \frac{d_k}{n_k^2}} = \sqrt{SE(\hat{H}_N(t_{i-1}))^2 + \frac{d_i}{n_i^2}}$

31) A 95% CI for θ is $\hat{\theta} \pm 1.96\, SE(\hat{\theta})$, but other CIs for θ exist. See p. 7.

For HW3

in complete (or partial) remission induced by the drug Prednisone were given the drug 6-MP (Klein and Moeschberger p2 and 186)

= time to relapse

	$t_{(j)}$	γ_j	t_i (to=0)	$\Sigma t_{(j)} \Sigma t_i$	n_i	d_i	$\hat{H}_N(t_i)$	$(SE(\hat{H}_N(t_i)))^2$
							0	
1	6	1	6		21	3	$0+\frac{3}{21}=\frac{3}{21}\approx .1428$	$0+\frac{3}{(21)^2}\approx .0068$
2	6	1						
3	6	1						
4	6	0						
5	7	1	7		17	1	$.1428+\frac{1}{17}=.2017$	$.0068+\frac{1}{(17)^2}\approx .0103$
6	9	0						
7	10	1	10		15	1	$.2017+\frac{1}{15}=.2683$	$.0103+\frac{1}{(15)^2}\approx .0147$
8	10	0						
9	11	0						
10	13	1	13		12	1	$.2683+\frac{1}{12}=.3517$	$.0147+\frac{1}{(12)^2}=.0217$
11	16	1	16		11	1	$.3517+\frac{1}{11}=.4426$	$.0217+\frac{1}{(11)^2}=.0299$
12	17	0						
13	19	0						
14	20	0						
15	22	1	22		7	1	$.4426+\frac{1}{7}=.5854$	$.0299+\frac{1}{(7)^2}=.1725$
16	23	1	23		6	1	$.5854+\frac{1}{6}=.7521$	$.2243+\frac{1}{(6)^2}=.2732$
17	25	0						
18	32	0						
19	32	0						
20	34	0						
21	35	0						

So a 95% CI for $H_y(13)$

is $\hat{H}_N(13) \pm 1.96\, SE(\hat{H}_N(13)) = (L, U)$

use $(\max(0, L), U)$

$= .3517 \pm 1.96\sqrt{.0217} = .3517 \pm .2888$

$= (.0630, .6404)$

↑ skip to 34) For HW3

32) * p? Consider the life table estimator with interval $I_j = [t_{j-1}, t_j)$, d_j and n_j'.

Then $\hat{h}_L(t) = \dfrac{d_j}{\left(n_j' - \frac{d_j}{2}\right)(t_j - t_{j-1})}$, $t_{j-1} \le t < t_j$.

but $\hat{h}_L(t)$ is not defined on the last interval $[t_{m-1}, \infty)$

Sometimes $t^* = \frac{t_{j-1}+t_j}{2}$ is used,

ex} HW 2 output

I_j	t^*	d_j	n_j'	$\hat{h}_L(t^*)$
$[0,50)$	25	16	66.5	$\frac{16}{(66.5-\frac{16}{2})(50-0)} = .0054$
$[50,100)$	75	11	49	$\frac{11}{(49-\frac{11}{2})(100-50)} = .005058$

33) * p? consider the KM estimator with

$t_0 = 0$

t_1	n_1	d_1
t_2	n_2	d_2
⋮		
t_m	n_m	d_m

or step function for $0 < t_j \le t < t_{j+1}$ but not defined on $(0, t_1)$ or (t_m, ∞).

Then $\hat{h}_{KM}(t_j) = \dfrac{d_j}{n_j(t_{j+1} - t_j)}$ for $j = 1, \ldots, m-1$.

p35 $\hat{h}_{KM}(t) = \hat{h}_N(t)$ ^ Nelson Aalen

ex}

t_i	n_i'	d_i	$\hat{h}_{KM}(t_i)$
10	18	1	$\frac{1}{18(19-10)} = .00617$
19	15	1	$\frac{1}{15(30-19)} = .00606$
30	13	1	

34) p40 $F_Y(t_Y(p)) = p$ so $S_Y(t_Y(p)) = 1-p$, $0 < p < 1$.

$t_Y(.5)$ is the median survival time.

$\hat{t}_Y(p)$ is found from $\hat{S}(t)$ with over and down lines

but computer output also gives 95% CI for $t_Y(p)$ when $p = .25$ .5 .75

P.X The CI has the form $\hat{t}(p) \pm 1.96\, SE(\hat{t}(p))$.

~~Also see p337, 338~~

ex) SAS Output

Quartile Estimates

Percent	Point Estimate	95% CI lower	95% CI upper
75	.	220.0	.
50	210.0	63.0	1296.0
25	63.0	18.0	195.0

So $\hat{t}_y(.5) = 210$, 95% CI for $t_y(p)$ is (63, 1296)

$\hat{t}_y(.25) = 63$ (18, 195)

If $\hat{S}(t) > 1-p \quad \forall t$, then $t_y(p)$ can't be estimated.

Hence $t_y(.75)$ could not be estimated in this ex. see Graph plots

35) P.53 $\mu = EY = \int_0^\infty y f(y)\,dy = \int_0^\infty t f(t)\,dt = \int_0^\infty S(t)\,dt$ (if $\lim_{t\to\infty} tS(t) \to 0$)

end exam 1 material

$\hat{\mu} = \sum_{j=1}^{n} \hat{S}(t_{(j-1)})(t_{(j)} - t_{(j-1)})$

tends to be biased if $\hat{S}(t_{(n)}) \neq 0$

eg if $t_{(n)}$ is censored

ch3 1) Regression is the study of the conditional distribution of $Y|\underline{x}$ where the vector of predictors $\underline{x} = (x_1, \ldots, x_p)$.

2) P.55 The predictors are also called independent variables, covariates, risk factors or explanatory variables.

3) P.55 The simplest example is $\underline{x} = (x_1) = x_1 = x$ (scalar)

where $x = \begin{cases} 1 & \text{new treatment} \\ 0 & \text{standard treatment or placebo = sham treatment} \end{cases}$

new is better than old

If you had enough data at $\underline{x} = \underline{x}_0$, say

$Y_1^*(\underline{x}_0), \ldots, Y_n^*(\underline{x}_0)$ you could make, for example, the KM estimator for various values of $\underline{x}_0$ and plot the survival curves eg $\hat{S}_{KM}(t|\underline{x}_1)$, $\hat{S}_{KM}(t|\underline{x}_2), \ldots, \hat{S}_{KM}(t|\underline{x}_j)$

ex see handout for 1989 US survival times for BM WM BF WF.

4) Problem: often for each distinct value of $\underline{X}$ you have only one $Y^*|\underline{x} = \min(Y, Z|\underline{x})$.

5) data $\left.\begin{matrix} Y_1^*, \delta_1, \underline{x}_1 \\ Y_2^*, \delta_2, \underline{x}_2 \\ \vdots \\ Y_n^*, \delta_n, \underline{x}_n \end{matrix}\right\}$ want to estimate

$h_i(t) = h(t|\underline{x}_i) = h_{Y_i|\underline{x}_i}(t)$.

$= h_{Y_i|\underline{\beta}'\underline{x}_i}(t)$.

6) ~~p57~~ The most general proportional hazards model is

$h_i(t) = \psi(\underline{\beta}'\underline{x}_i)\, h_0(t)$,

$\psi(w) \geq 0$ and $\psi(0) = 1$.

7) ~~p58~~ Know for final | The (standard) Cox regression model or proportional hazards model is

$h_i(t) = h_{Y_i|\underline{x}_i}(t) = h_{Y_i|\underline{\beta}'\underline{x}_i}(t) = e^{\underline{\beta}'\underline{x}_i} h_0(t)$.

8) ~~p58~~ The sufficient Predictor (SP) = linear component =

linear predictor = $\eta_j = \sum_{i=1}^{p} \beta_i x_{ij}$

$= \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{ip}$.

9) ~~P150~~ $h_{y|sp}(t) = e^{sp} h_0(t)$ and $\frac{h_{y|sp}(t)}{h_0(t)} = e^{sp} = \exp(\beta' x)$

SO $sp = \log\left(\frac{h_{y|sp}(t)}{h_0(t)}\right)$

10)* the estimated sufficient predictor $ESP = \hat{\beta}' x$.

11) ~~298, 103~~ P.15, know for final Given $\hat{\beta}$ from output and x, be able to find $ESP = \hat{\beta}' x$ and $e^{\hat{\beta}' x} = \hat{h}_i(t)$

ex) Heart transplant data Allison p120

Y = days from acceptance until death

$y^* = surv = \min(Y, z)$ = possible censored survival time

$x_1 = trans = \begin{cases} 1 & \text{received heart transplant} \\ 0 & \text{else} \end{cases}$

$surg = \begin{cases} 1 & \text{transplant before date of acceptance} \\ 0 & \text{else} \end{cases}$

ageaccpt = age at date of acceptance

	variable	df	estimate	SE	Wald chi-square $\chi^2_{obs} = z_{obs}^2$	pval Pr > chisq	risk ratio
x_1	Trans	1	$\hat{\beta}_1$ −1.70814	.2786	37.59	.0001	.181
x_2	surg	1	$\hat{\beta}_2$ −0.42140	.3710	1.29	.2560	.656
x_3	ageaccpt	1	$\hat{\beta}_p$ 0.05486	.0151	15.16	.0001	1.060

a) Find the ESP $\hat{\beta}' x$ if $x = (1, 0, 64)$.

Soln a) $ESP = \hat{\beta}'\underline{x} = -1.70814(1) - 0.4214(0) + .05861(64)$

$= 2.0429$

b) Find $\hat{h}_i(t) = e^{\hat{\beta}'\underline{x}}\hat{h}_0(t) = e^{2.0429}\hat{h}_0(t)$

$= 7.7129\,\hat{h}_0(t)$.

If $\underline{x}=\underline{0}$ makes sense, $\underline{x} = (1, 0, 64)$ is about 7.7129 times more hazardous.

~~§3.4~~ §3.3

12) know for final ~~p 69~~ p101

A 95% (Wald) CI for β_i is $\hat{\beta}_i \pm 1.96\, SE(\hat{\beta}_i)$

ex) A 95% CI for β_2 is

$\hat{\beta}_2 \pm 1.96\, SE(\hat{\beta}_2) = -.4214 \pm 1.96(.3710)$

$= -.4214 \pm .72716 = (-1.14856, .30576)$

reasonable values for β_2

CI includes 0 so x_2 may not be important given that x_1 and x_3 are in the model.

13) ~~p69-70~~ p99 know for final Test whether x_j is needed in the model given that the other predictors $x_1, \ldots, x_{j-1}, x_{j+1}, \ldots, x_p$ are in the model.

4 step (Wald) test i) $H_0: \beta_j = 0$ t/A $\beta_j \neq 0$

ii) $z_{0j} = \dfrac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$ or $x^2_{0j} = z^2_{0j}$ often from output

use $\delta = .05$ if δ is not given

iii) pvalue $= 2P(Z < -|z_{0j}|) = P(\chi^2_1 > x^2_{0j})$ often from output

iv) reject H_0 if pval $< \delta$ and conclude x_j is needed in the model given the other predictors are in the model

fail to reject H_0 if pval $> \delta$ and conclude x_j is not needed in the model given

ex) a) perform a 4 step test for $\beta_2 = 0$ without output

b) $\beta_3 = 0$ with output

Soln a) i) H_0 $\beta_2 = 0$ H_A $\beta_2 \neq 0$

ii) $z_{02} = \frac{-.4214}{.3710} = -1.136$

(output 1.29 $= z_{02}^2$)

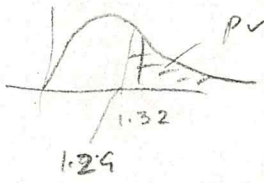
round To to 2 digits

iii) pval $= 2P(z < -|z_{02}|) = 2P(z < -1.14)$

$= 2(.1271) = .2542$

(output pval = .256)

df	.250
1	1.32



iv) pval $> \delta = .05$ so fail to reject H_0

Surg is not needed in the survival model given that Trans and age accpt are in the model.

b) i) H_0 $\beta_3 = 0$ H_A $\beta_3 \neq 0$

ii) $X_{02}^2 = 15.16$

iii) pval $= .0001 < \delta = .05$

iv) reject H_0, age accpt is needed in the survival model given Trans and Surg are in the model.

14) If $-|z_{03}| \leq -3.9$, take pvalue $= 0$ from the normal tab

15) The null model $h_i(t) \equiv h_0(t)$ has no predictors.

16) The full model has all of the predictors $\underline{x}$ in the model

17) Know for final (p103, 98) The partial likelihood ratio test PLRT is used to test whether $Y \perp\!\!\!\perp \underline{x}$, that is whether the survival times are independent of the predictors.

output variables in model	$-2 \log \hat{L}$
none	$-2 \log \hat{L}$ (none)
$x_1, \ldots, x_p$	$-2 \log \hat{L}$ (Full)

or Model Fit Statistics

Criterion	Without Covariates	With Covariates	Model Chi-Sq	DF	Pr > ChiSq
Likelihood Ratio or $-2 \log L$	$-2 \log \hat{L}$ (none)	$-2 \log \hat{L}$ (full)	$X^2(NIF)$	p	$P(\chi^2_p > X^2(NIF))$

Testing Global Null Hypothesis: BETA=0

Test	Chi-Sq	DF	Pr > ChiSq
Likelihood Ratio	$X^2(NIF)$	p	$P(\chi^2_p > X^2(NIF))$

p342

4 step PLRT i) H_0 $\underline{\beta} = 0$ H_A $\underline{\beta} \neq 0$

ii) $X^2(NIF) = (-2 \log \hat{L}(\text{none})) - (-2 \log \hat{L}(\text{full}))$

often from output

iii) $pval = P(\chi^2_p > X^2(NIF))$

iv) If $pval < \delta$, reject H_0 and conclude that there is a PH survival relationship between Y and $\underline{x}$. If $pval \geq \delta$ fail to reject H_0 and conclude that there is not a PH survival relationship between Y and X.

x_1, x_2, x_3, x_4, x_5 $-2\log L = 162.479$

none $-2\log L = 177.667$

or output R: likelihood ratio test = 15.188 on 5 df $p = .00959$

SAS Testing Global Null hypothesis Beta = 0

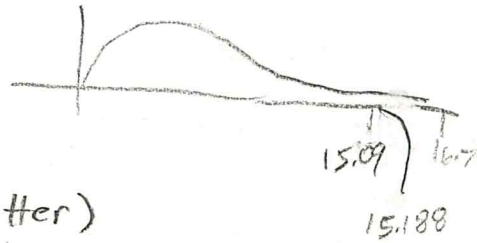
Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	15.188	5	.00959

4 step test i) H_0 $\beta = 0$ H_A $\beta \neq 0$

ii) $\chi^2(NF) = 177.667 - 162.479 = 15.188$

iii) $pval = P(\chi^2_5 > 15.188)$

	.010	.005
5	15.09	16.75



$.005 < pval < .010$
(output .00959 is better)

iv) reject H_0 there is a PH survival relationship between survival times Y and the predictors x_1, x_2, x_3, x_4, and x_5.

ex) Treatment A and B for leukemia patients in remission. Time in weeks until relapse

$x = \sum_{1}^{0}$ trt A / trt B

$p = 1$ so $\beta = \beta$ (a scalar)

R likelihood ratio test = 1.32 on 1 df, $p = 0.25$, $n = 46$

i) H_0 $\beta = 0$ H_A $\beta = 0$

ii) $\chi^2(NF) = 1.32$

iii) $pval = .25$

iv) fail to reject H_0 there is not a PH survival relationship between relapse times and X (so no difference between treatments A and B)

18) * Assume $sp = x_1\beta_1 + \cdots + x_p\beta_p = \underline{\beta}'\underline{x} =$

$\underline{\beta}_R'\underline{x}_R + \underline{\beta}_0'\underline{x}_0$ is a useful full model.

Let the reduced model $\beta_{R1}x_{R1} + \cdots + \beta_{Rr}x_{Rr}$

$= \underline{\beta}_R'\underline{x}_R$.

19) know for final Assume the full model is useful.
The change in partial likelihood ratio test is used to test whether the reduced model can be used instead of the full model, that is, whether $\underline{x}_0$ is needed in the model given $\underline{x}_R$ is in the model. Fit the full and reduced models to get $x^2(M|F)$ and $x^2(M|R)$ (used to test $\underline{\beta}=\underline{0}$ and $\underline{\beta}_R=$

20) Notice that $x^2(R|F) \equiv x^2(M|F) - x^2(M|R) =$
$[-2\log\hat{L}(\text{none})] - [-2\log\hat{L}(\text{full})] - (-2\log\hat{L}(\text{none})] - [-2\log\hat{L}(\text{red})])$
$= [-2\log\hat{L}(\text{red})] - [-2\log\hat{L}(\text{full})] = -2\log\left[\frac{\hat{L}(\text{red})}{\hat{L}(\text{full})}\right]$

21) ~~p76~~ p109 know for final 4 step change in PLR test

i) H_0 reduced model is good H_A use the full model

ii) $x^2(R|F) = x^2(M|F) - x^2(M|R)$

iii) pval $= P(\chi^2_{p-r} > x^2(R|F))$

iv) reject H_0 if pval $\leq \delta$ and conclude the full model should be used
fail to reject H_0 if pval $\geq \delta$ and conclude the reduced model is good

ex) AIDS

received placebo, low dose, or high dose

$V1 = \begin{cases} 1 & \text{low dose} \\ 0 & \text{else} \end{cases}$

$V2 = \begin{cases} 1 & \text{high dose} \\ 0 & \text{else} \end{cases}$

Smith p179-180

time until blood test was positive, blood test taken each day for a month

note: $(V1, V2) = (0,0) \rightarrow$ placebo

	R coef	SAS parameter estimate	se coef Standard error	z	chisquare	p Pr > chisq
V1	−1.51		.528	−2.86	8.1796	.0043
V2	−1.03		.455	−2.26	5.1076	.0240

R: likelihood ratio test = 8.99 on 2 df, $p = 0.0111$

SAS Test	chi sq	df	Pr > chi sq
Likelihood ratio	8.99	2	.0111

Let the reduced model have V1 alone

R: likelihood ratio test = 3.88 on 1 df p=

SAS Test	chisq	df	pr > chisq
Likelihood ratio	3.88	1	

Test whether the reduced model is good

i) H_0: the reduced model is good H_A use the full model

ii) $X^2(RIF) = X^2(WIF) - X^2(WIR) = 8.99 - 3.88 = 5.11$

iii) $pval = P(\chi^2_{2-1} > 5.11)$

df	.025	.01
1	5.02	6.03

$.01 < pval < .025$

iv) reject H_0, use the full model

ex ~~p79~~ Variable in model $-2\log L$

reduced A_2, A_3, N 165.508

full A_2, A_3, N, A_2N, A_3N 162.479

Do a 4 step test.

i) H_0 the reduced model is good H_A use the full model

ii) $X^2(R|F) = -2\log L(R) - [-2\log L(F)]$

$= 165.508 - 162.479 = 3.029$

iii) $p - r = 5 - 3 = 2$ = # terms left out of the full model

$pval = P(\chi^2_2 > 3.029)$

df	.25	.1
2	2.77	4.61

$.1 < pval < .25$

iv) fail to reject H_0, the reduced model is good.

See notes 43 for a crucial point.

~~§3.2~~ §4.2

22) ~~*p59~~ A variate is a variable that takes on numerical values. A factor A is a variable that takes on a categories, called levels.

ex) gender $< \begin{matrix} M \\ F \end{matrix}$ $a = 2$

ex) tumor type $<$ squamous, adeno, small cell $a = 3$

23) p59 Let $x_{jA} = \begin{cases} 1 & A=j \\ 0 & \text{else} \end{cases}$ $j = 1, \ldots, a$.

$x_{1A}, \underbrace{x_{2A}, \ldots, x_{aA}}$

To include a factor into the model, add the $a-1$ indicator variables $x_{2A}, \ldots, x_{aA}$ to the model. x_{1A} is not added to the model. A factor has $a-1$ degrees of freedom.

ex) $x_{1A} = \begin{cases} 1, & \text{squamous} \\ 0, & \text{else} \end{cases}$ $x_{2A} = \begin{cases} 1, & \text{adeno} \\ 0, & \text{else} \end{cases}$ $x_{3A} = \begin{cases} 1 & \text{small cell} \\ 0 & \text{else} \end{cases}$

$(x_{2A}, x_{3A}) = \begin{cases} (0,0) & \rightarrow \text{squamous} \\ (1,0) & \rightarrow \text{adeno} \\ (0,1) & \rightarrow \text{small cell} \end{cases}$

In general $(x_{2A}, \ldots, x_{aA}) = (0, \ldots, 0)$ or has a 1 in the jth position. (= $j+1$th category)

24) p60 The x_j corresponding to variates or the indicator variables from factors are called <u>main effects</u>.

An <u>interaction</u> is a product of 2 or more main effects. Convention: the interaction between a variate x_1 and a factor A with indicator variables $x_{2A}, \ldots, x_{aA}$ is incorporated into the model with is $x_1 x_{2A}, x_1 x_{3A}, \ldots, x_1 x_{aA}$

An interaction between factor A and factor B with indicators $x_{2B}, \ldots, x_{bB}$ is incorporated

into the model with the $(a-1)(b-1)$ pairs $x_{2A}x_{2B}$ ... $x_{2A}x_{bB}$

$\vdots$

$x_{aA}x_{2B}$ ... $x_{aA}x_{bB}$.

1st order interaction

2nd order interaction

2) Sometimes use $x_{12} = x_1 x_2$, $x_{123} = x_1 x_2 x_3$ etc.

ex) p133 Suppose x_1 is a variate and $x_2 = \begin{cases} 1 & A = C_2 \\ 0 & A = C_1 \end{cases}$

where A takes on 2 levels C_1 and C_2.

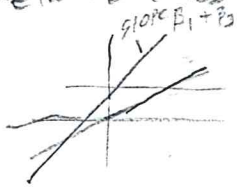
A first order model with interaction is

$$SP = \beta_0 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2.$$

If $x_2 = 1$, $SP = \beta_2 + (\beta_1 + \beta_3) x_1$

If $x_2 = 0$, $SP = \beta_0 x_1$

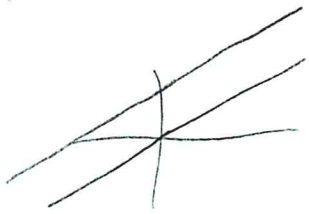
} 2 unrelated lines



If $\beta_3 = 0$, there are 2 parallel lines!

$x_2 = 1$: $SP = \beta_2 + \beta_1 x_1$

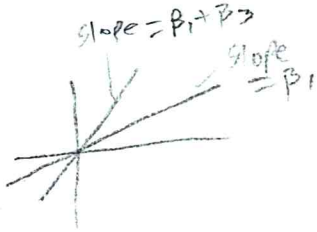
$x_2 = 0$: $SP = \beta_1 x_1$.



If $\beta_2 = 0$, the 2 lines both have intercept at the origin.

$x_2 = 1$: $SP = (\beta_1 + \beta_3) x_1$

$x_2 = 0$: $SP = \beta_1 x_1$



If $\beta_2 = \beta_3 = 0$, the 2 lines are coincident

$SP = \beta_1 x_1$

2b) ~~p60~~ Convention if an interaction is in the model, also include the corresponding main effects

ex) If $x_1 x_2$ is in the model, x_1 and x_2 should be in the model.

27) Models with interactions get complex rapidly. Sometimes one fits a full model with 1st order interactions and a reduced model without the interactions in the hope that the reduced model will be good.

~~§3.3 (stim)~~

28) ~~p63~~ The baseline hazard function $h_0(t)$ does not need to be known to obtain $\hat{\beta}$, and $\hat{h}_0(t)$ can be obtained after finding $\hat{\beta}$.

29) ~~p66~~ $\hat{\beta}$ is a maximum partial likelihood estimate MPLE and estimation is similar to that of MLEs.

~~§3.5~~ §3.5 30) A model for variable selection is $SP = \underset{\sim}{\beta}^T \underset{\sim}{x} = \underset{\sim}{\beta}_S^T \underset{\sim}{x}_S + \underset{\sim}{\beta}_E^T \underset{\sim}{x}_E = \underset{\sim}{\beta}_S^T \underset{\sim}{x}_S$ where

$\underset{\sim}{x} = (\underset{\sim}{x}_S^T, \underset{\sim}{x}_E^T)^T$ is a $p \times 1$ vector of predictors

$\underset{\sim}{x}_S$ is an $r_S \times 1$ vector and $\underset{\sim}{x}_E$ is a $(p - r_S) \times 1$ vector.

Given $\underset{\sim}{x}_S$ is in the model, $\underset{\sim}{\beta}_E = \underset{\sim}{0}$ and E denotes the subset of variables that can be eliminated given the subset S is in the model.

31) $SP = \underset{\sim}{\beta}^T \underset{\sim}{x}$ is the full model

$SP = \underset{\sim}{\beta}_I^T \underset{\sim}{x}_I$ is a submodel.

32) * The full model is also a submodel.

33) Suppose S is a subset of I.

Then $SP = \underset{\sim}{\beta}^T \underset{\sim}{x} = \underset{\sim}{\beta}_S^T \underset{\sim}{x}_S = \underset{\sim}{\beta}_S^T \underset{\sim}{x}_S + \underset{\sim}{\beta}_{(I|S)}^T \underset{\sim}{x}_{I|S} + \underset{\sim}{0}^T \underset{\sim}{x}_0$

$= \underset{\sim}{\beta}_I^T \underset{\sim}{x}_I$ where $\underset{\sim}{x}_{(I|S)}$ denotes the predictors in I but not in S. Hence $\underset{\sim}{\beta}_0 = \underset{\sim}{0}$ if $S \subseteq I$.

Hence $\text{corr}(\underset{\sim}{\beta}^T \underset{\sim}{x}, \underset{\sim}{\beta}_I^T \underset{\sim}{x}) = 1$ if $S \subseteq I$.

34) An EE plot is a plot of $ESP(I)$ vs ESP where $ESP(I) = \hat{\underset{\sim}{\beta}}_I^T \underset{\sim}{x}_I$ and $ESP = \hat{\underset{\sim}{\beta}}^T \underset{\sim}{x}$.

35) ~~pgl~~ SAS uses $AIC(I) = -2 \log \hat{L}(I) + 2 r_I$ where $\hat{\underset{\sim}{\beta}}_I$ is $r_I \times 1$.

36) ~~pgl~~ Rule 1 If an interaction (x_3 x_7 x_9) is in the model, the main effects (x_3, x_7 and x_9) should also be in the model.

37) Rule 2 If $x_{i+1}, x_{i+2}, \ldots, x_{i+a-1}$ are the $a-1$ indicator variables corresponding to factor A, either keep all of them in the model or delete all of them.

38) [Assume that the full model] is useful and !!! that all of the variables are initially equally important. (This assumption is violated if Factor A corresponds to a treatments. Then $x_1, x_2, \ldots, x_{a-1}$ corresponding to factor A should be included in all submodels.)

39) It is desirable to eliminate expensive predictors and predictors such as interactions that make the model hard to understand.

40) It is desirable to have a full and reduced model before collecting data. Variable selection is a form of data snooping and often must be done to get a reasonable model. But after variable selection, the model is too good to be true. pvalues for $H_0: \beta_i = 0$ are too small and if x_I is the selected model, then the pvalue for the change in PLRT that uses x_I as the reduced model is too high.

41) If an important predictor is deleted, the PH model may no longer hold. If you leave in predictors that should be deleted, then you are "fitting noise" and the model may appear to be

better than it is.

42) Numerical methods

i) want $-2\log \hat{L}(\hat{\beta}_I)$ small but $-2\log \hat{L}(\hat{\beta}_{full}) \leq -2\log \hat{L}(\hat{\beta}_I)$ since adding predictors does not increase and usually decreased $-2\log \hat{L}$.

ii) want $n \geq 5 r_I$ and preferably $n \geq 10 r_I$ where r_I = # variables in submodel I.

iii) Do not use more predictors than the min AIC model I_{min}.

iv) Start with the model I_0 with fewest predictors such that $AIC(I_0) \leq AIC(I_{min}) + 2$. Also look at all models I_j with fewer predictors than I_0 such that $AIC(I_j) \leq AIC(I_{min}) + 7$.

v) want $corr(ESP(I), ESP) \geq 0.95$. So the EE plot should cluster tightly about the identity line with unit slope and 0 intercept for models I_j with $r_{I_j} \leq r_{I_{min}}$) let

43) $\Delta(I) = AIC(I) - AIC(I_{min})$. Models with $\Delta(I) \leq 2$ are good,

$2 < \Delta(I) \leq 7$ are borderline

and models with $\Delta(I) > 7$ should not be used.

44) All subsets Compute all $2^p - 1$ submodels

§5.4

I that contain 1 or more X_i. Compute AIC(I) for each model and keep track of 1 to 5 models for each subset size $r_I = 1, 2, \ldots, p$. This method is impractical for $p \geq 10$ or so.

45) * ~~p82~~ backward elimination

At step m there is a pool of $p-m+1$ variables

Delete the predictor whose omission results in the smallest increase of $-2\log\hat{L}$ (or in the smallest $AIC(I_{p-m+1})$ or omit the variable with the largest pvalue), say x_j^*, $j=p$

step m=1: start with the full model, delete x_p^*

m=2: p-1 variables left, delete x_{p-1}^*

m=3: p-2 variables left, delete x_{p-2}^*

⋮

m=p-1, 2 variables left, delete $x_{p-(p-2)}^*$

ex Suppose x_1 and x_2 are left

$-2\log\hat{L}(\hat{\beta}_I)$ $AIC = -2\log\hat{L} + 2r_I$

x_1	1000	1002
x_2	2000	2002

Delete x_2 and $I_1 = \{x_1\}$

46*) Forward selection

After step m, I_m contains m predictors. Add the predictor that decreases $-2\log\hat{L}$ the most (or that makes AIC(I_m) the smallest or ~~add~~ the predictor that would have the ~~smallest~~ p-value)

step 1	add x_1^*	$I_1 = \{x_1^*\}$
2	add x_2^*	$I_2 = \{x_1^*, x_2^*\}$
⋮		
p-1	add x_{p-1}^*	$I_{p-1} = \{x_1^*, \dots, x_{p-1}^*\}$
p	add x_p^*	$I_p = \{x_1, \dots, x_p\}$ = full model

47) Forward selection and backward elimination each result in p models

$\{x_1^*\}, \{x_1^*, x_2^*\}, \dots, \{x_1^*, \dots, x_p^*\}$

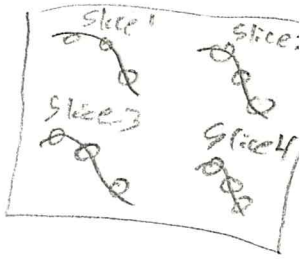
Look for the model with the smallest AIC I_{min} then the model I_0 with the fewest predictors such that $AIC(I_0) \leq AIC(I_{min}) + 2$.

Backwards: an increase in $-2\log\hat{L}$ of more than 4 may be troubling. The increases may be small and then suddenly large (>10). A large increase suggests that an important predictor has been omitted.

Forwards: A large decrease (>10) suggests that an important predictor has been included.

48) Stepwise selection is a modification of forward selection or backward elimination.

make an EE plot and slice the ESP.
Make coxPH $\hat{S}_{\underline{x}}(t)$ for $\underline{x}$ in each slice and add points of $\hat{S}_{KM}$ computed from each slice.
If submodel I is good, the plotted points should track the $\hat{S}_{\underline{x}}$ for each slice.



Look at pvalues in the Wald test to see if any predictors can be deleted.

50) For a given number of variables r, look at the model from forward selection or backward elimination with the smallest $-2\log\hat{L}$ which is equivalent to the model with the smallest $AIC = -2\log\hat{L} + 2r$.

51) For factors and interactions, predictors can have more than 1 df.

52) Assume the full model is good and all predictors equally important. Rules of thumb for a good PH submodel I in roughly decreasing order of importance.

i) Do not use more predictors than the min AIC model I_{min}.

ii) plot from 49) for I looks like plot for full model

iii) $corr(ESP, ESP(I)) \geq 0.95$ where $ESP = \hat{\beta}'\underline{x}$ $ESP(I) = \hat{\beta}_I'\underline{x}_I$

iv) The plotted points in the EE plot cluster tightly about the identity line.

v) want pvalue ≥ 0.01 for the change in PLR test that uses I as the reduced model (for variable selection use $\delta = .01$ instead of $\delta = .05$).

vi) want $r_I \leq n/10$.

vii) want $-2\log \hat{L}(\hat{\beta}_I) \geq -2\log \hat{L}(\hat{\beta}_{FULL})$ but close.

viii) want $AIC(I) \leq AIC(I_{min}) + 7$ where I_{min} is the min AIC model found with the variable selection procedure.

ix) want hardly any predictors with pvalues $> .05$

x) want few predictors with pvalues between .01 and .05

But for factors with at indicators modify ix) and x) so that the indicator with the smallest pvalue is examined.

5.3) ~~p65~~ Now assume there is at least one predictor variable, eg treatment, that must be in the PH model.

Strategy i) keep these variables in the model but do forward or backward elimination on the remaining variables

ii) omit the important variables, do variable selection on the remaining variables, then add the important variables to the model

>T) know for proj given a list of variables with their $AIC = -2\log L + 2r$ be able to find the submodel I_0 that has the fewest predictors r_0 and smallest AIC such that $AIC(I_0) \leq AIC(I_{min}) + 2$.

Other candidates I have fewer predictors than I_0 and $AIC(I) \leq AIC(I_{min}) + 7$

ex ~~PSG~~ Prostate cancer

variables	$-2\log L$	$AIC = -2\log L + 2r$	
none	36.349	36.349	
age	36.269	38.269	
shb	36.196	38.196	
size	29.042	31.042	← candidate
index	29.127	31.127	
age, shb	36.151	40.151	
age, size	28.854	32.854	
age, index	28.760	32.760	
shb, size	29.019	33.019	
shb, index	27.981	31.981	
size, index	23.533	27.533	← $I_{min} = I_0$
age, shb, size	28.852	34.852	
age, shb, index	27.893	33.893	
age, size, index	23.269	29.269	
shb, size, index	23.508	29.508	
age, shb, size index	23.231	31.231	
Full			

55) know for final given summaries on several models, be able to pick out the "best starting model."

	M_1	M_2	M_3 = min AIC	M_4
ex)	Full	age, size, index	size index	size

age, shb size index

	M1	M2	M3	M4
# predictors	4	3	2	1
# with .01 < Wald pval < .05	1	2	1	0
# with Wald pval > .05	2	1	0	0
$-2 \log L$	23.231	23.269	23.533	29.042
AIC	31.231	29.269	27.533	31.042
pval for change in PLR test	1.0	.8454	.8598	.121
		(.038, 1)	(.302, 2)	(5.811, 3)

M1 and M2 have more predictors than the min AIC model.

M4's AIC is too large to be the starting model

so use M3 as the starting model.

If M4 had $-2\log L = 27.042$, AIC = 29.042 pval = .283

then M4 would be the starting model.

Any model with pval < .01 has pval too small

end exam 2 material

56) ~~p 57-59~~ know for final

log rule: if X is a nonnegative variable

and $\frac{\max X}{\min X} > 10$, use $\log X$ instead of X

in the model

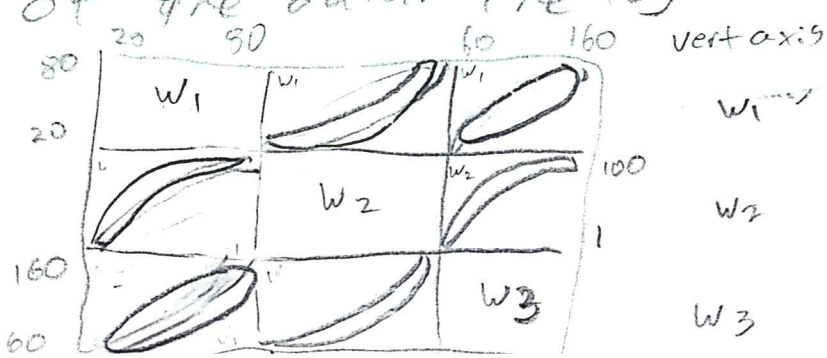
57) It is often useful to transform variables

$X = t(w)$ to remove non linearities from a

scatterplot matrix of the data. The log

rule is useful.

ex) log (W2) makes sense



see 9.11.4, 16.1

Let $h_i(t) = h_{\underline{x}}(t) = e^{\beta_i} h_0(t)$

$= \exp(\beta_1 x_1 + \cdots + \beta_i x_i + \cdots + \beta_p x_p)\, h_0(t)$,

so $e^{\underline{x}'\underline{\beta}} = \frac{h_{\underline{x}}(t)}{h_0(t)}$ and $SP = \underline{x}'\underline{\beta} = \log\left(\frac{h_{\underline{x}}(t)}{h_0(t)}\right)$

Suppose x_i is increased by r units while the other x_j are held fixed.

$SP(x_i + r) = \beta_1 x_1 + \cdots + \beta_i(x_i + r) + \cdots + \beta_p x_p = SP + r\beta_i$

$h_{i|x_i+r}(t) = \exp(r\beta_i)\exp(\underline{\beta}'\underline{x})\, h_0(t) = \exp(r\beta_i)\, h_i(t)$

so the hazard ratio $\frac{h_{i|x_i+r}(t)}{h_0(t)} = \exp(r\beta_i)\frac{h_i(t)}{h_0(t)}$

changes by a factor of $\exp(r\beta_i)$. The log hazard ratio $\log\frac{h_{i|x_i+r}(t)}{h_0(t)} = r\beta_i + \log\frac{h_i(t)}{h_0(t)} = r\beta_i + \underline{x}'\underline{\beta}$

So β_i is the change in log hazard ratio when x_i is changed by 1 unit with all other x_j held fixed.

may not be possible, eg $x_1 = x, x_2 =$

factor makes sense if $x_1 = \cdots = x_{a-1} = 0$ but not if one = 1

{ factor with levels $x_1, \ldots, x_{a-1}, a$ zz, if $x_1 = 0$ and $x_2 = 1$ to change x_1 to $x_1 + 1 = 1$ need to change x_2 to $x_2 - 1 =$

If $x_j = 0$ for $j = 1, \ldots, p$, then $h_i(t) = h_0(t)$ and changing x_i by 1 unit makes the hazard ratio $h_{i|x_i+1}(t) = e^{\beta}\frac{h_0(t)}{h_0(t)} = e^{\beta}$

R output includes $\exp(\hat{\beta}_i) = e^{\hat{\beta}_i}$

~~§3.3 Only read ex 3.12 on p 103–106, but~~

59} $\hat{h}_i(t) = \exp(\hat{\underline{\beta}}' \underline{x}_i)\, \hat{h}_0(t)$

$\hat{H}_i(t) = \exp(\hat{\underline{\beta}}' \underline{x}_i)\, \hat{H}_0(t)$

$\hat{S}_i(t) = [\hat{S}_0(t)]^{\exp(\hat{\underline{\beta}}' \underline{x}_i)}$

"forward" does not seem to work

~~Skip §3.9~~ 60] varsel in R: library(survival); library(MASS)
out <- coxph( ) ; stepAIC(out, direction = "backward")

Ch6 ~~Ch4~~ 1) ~~p111~~ Residuals are quantities calculated for each individual, and their behavior is roughly known when the fitted model is satisfactory.

~~Skip §4.1~~

2) ~~p112~~ The Cox Snell residual $r_{ci} = \exp(\hat{\underline{\beta}}' \underline{x}_i)\, \hat{H}_0(t_i)$

$= \hat{H}_{\underline{x}_i}(t_i)$, $i = 1, \ldots, n$ t_i the observed death or censoring time of obs i.

3) ~~p113~~ $-\log S(t) \sim EXP(1)$. So if $\hat{S}(t)$ is a good approx to $S(t)$, then $-\log \hat{S}(t_i) = \hat{H}_{\underline{x}_i}(t_i) = r_{ci}$ should behave like n obs's from a censored EXP(1) distribution.

4) ~~p115~~ p163 Let $\delta_i = \begin{cases} 1 & t_i \text{ uncensored} \\ 0 & t_i \text{ censored} \end{cases}$

then the martingale residual $r_{mi} = \delta_i - r_{ci}$

has mean 0 for uncensored cases.

$-\infty < r_{mi} \le 1$ and $r_{mi} < 0$ if $\delta_i = 0$ ie if case i is censored. ~~$\sum_{i=1}^{n} r_{mi} = 0$.~~

$\hat{H}(t) = -\log \hat{S}(t) \approx t$ (Exp(1)). So a cumulative hazard plot of t vs $\hat{H}(t)$ should follow the identity line.

Compute the KM estimators $\hat{S}$ of the r_{Ci} and plot r_{Ci} vs $\hat{H}(r_{Ci}) = -\log \hat{S}(r_{Ci})$.

This is a cumulative hazard plot of the residuals and should follow the identity line if the PH model is satisfactory. Departures from the identity line suggest that the PH model needs to be modified.

(p.123 $\hat{S}$ and $\hat{H}$ are not computed for censored cases. Recall that the KM step function $\hat{S}$ only jumps at uncensored cases.)

6) p123 saves the plots based on 5) are not very useful.

7) p123 plot ESP vs deviance residuals to find cases that are not well fitted by the PH model.

8) p123 Find the martingale residuals from the null model with no predictors. Plot x_i vs these residuals. If lowess curve is a straight line x_i can be used in the model instead of $t(x_i)$.

research!

Plot $t_\lambda(x_i)$ vs the residuals until the lowess curve is linear then $t_\lambda(x_i)$ should be used, eg for $t_\lambda(x_i) = \begin{cases} \frac{x_i^\lambda - 1}{\lambda}, & \lambda \neq 0 \\ \log x_i, & \lambda = 0 \end{cases}$

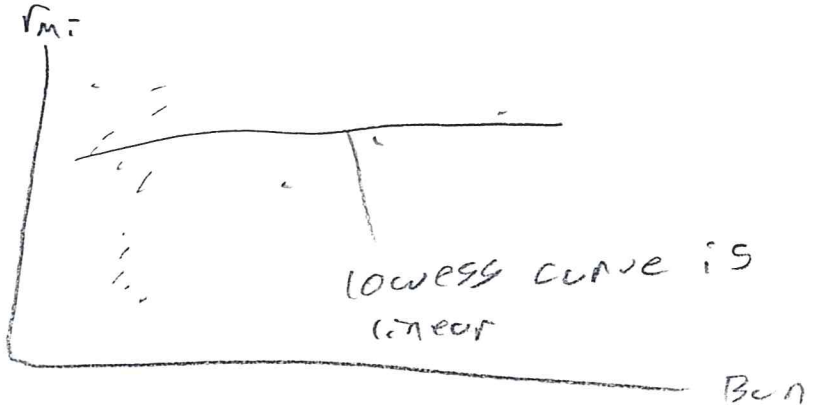
usually use $\lambda \in \{0, \pm\frac{1}{3}, \pm\frac{1}{2}, \pm 1, \pm 2, \pm 3\}$

9) p163 Suppose model contains $x_1, \ldots, x_p$. For $j = 1, \ldots, p$

leave out x_j, find the martingale residuals $r_{M(j)}$ and plot x_j vs $r_{M(j)}$. The lowess curve should be linear if x_j has the correct functional form. Otherwise plot $t_2(x_j)$ vs $r_{M(j)}$ until lowess curve in plot is linear.

research! (spss p 266)

ex p 130 Fig 4.8



log rule suggests using log Bun, but then lowess is curved so $(\log \text{bun})^2$ is needed.

§4.3 10) p 132 If $\underline{X}'\hat{\underline{\beta}}$ or the estimated hazard changes a lot if one case is omitted, that case is influential. Cases that change $\hat{\beta}_i$ for important predictors (such as treatment) are also important.

§4.4 p 132

11) $H_{\underset{\sim}{x}_i}(t) = \exp(\underset{\sim}{\beta}'\underset{\sim}{x}_i) H_0(t)$ so

$\log H_{\underset{\sim}{x}_i}(t) = \underset{\sim}{\beta}'\underset{\sim}{x}_i + \log H_0(t)$ or

$\log H_{\underset{\sim}{x}_i}(t) - \log H_{\underset{\sim}{x}_j}(t) = \underset{\sim}{\beta}'(\underline{x}_i - \underline{x}_j)$ is free of t.

12) Group data (eg slice $\hat{\underset{\sim}{\beta}}'\underset{\sim}{x}_i$) and compute $\hat{S}_{KMj}(t)$ for each group $j = 1, \ldots, J$.

$\hat{H}_j(t) = -\log \hat{S}_{KMj}(t)$. Plot $\log t$ vs $\log \hat{H}_j$ to form the log cumulative hazard plot.

If the PH model is a good approx to the data

and if each group has many deaths, then the plot should consist of roughly parallel curves. From p143, this plot does not seem to work very well if the data is grouped by levels of 1 variable.

13) The hazard ratio

$$\frac{h(\underline{x}_i, t)}{h(\underline{x}_j, t)} = \frac{\exp(\underline{\beta}'\underline{x}_i)\, h_0(t)}{\exp(\underline{\beta}'\underline{x}_j)\, h_0(t)} = e^{\underline{\beta}'(\underline{x}_i - \underline{x}_j)}$$

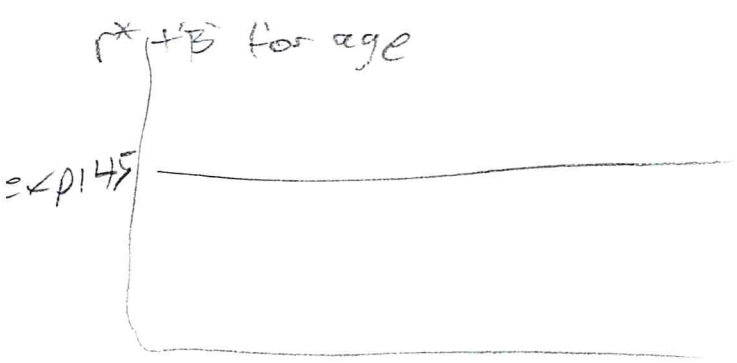
$\frac{e^a}{e^b} = e^{a-b}$

is free of t if the PH model is a good approx to the data. If the hazard ratio does depend on time, then the proportional hazards assumption is violated. This could occur if an explanatory variable depends on time.

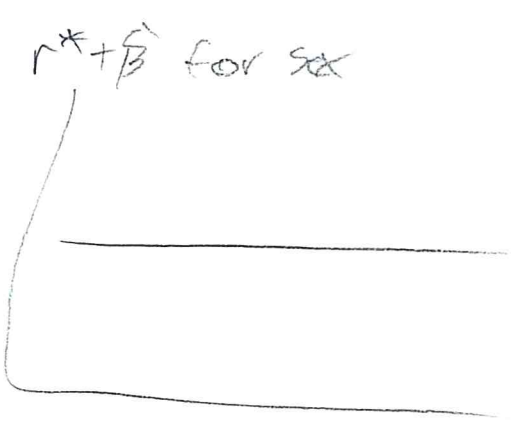
14) * p144 p148 The Schoenfeld residuals r^*_{pji} for the jth variable x_j satisfy

$$E(r^*_{pji}) \approx \beta_j(t_i) - \hat{\beta}_j \quad \text{for } i = 1, \ldots, n \text{ where}$$

$\beta_j(t_i)$ is the value of the coefficient at the ith death time t_i. Plot t_i vs $r^*_{pji} + \hat{\beta}_j$. If the lowess (or loess) curve is approximately horizontal, then the PH assumption is reasonable. Also, fit a line to the plot and test whether the slope of the line $= 0$. The p plots and tests are implemented in R with the cox.zph function



pval for slope = 0 is .391



pval for slope = 0 is .694

15) p148 compute $\hat{\beta}$ and for each variable x_j:

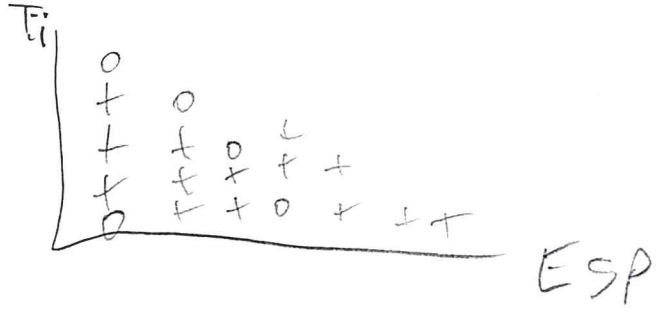
fit the model $SP = \underline{\beta}'\underline{x} + \gamma_j x_j t$. Test whether $\gamma_j = 0$. If the pvalues are $> .05$ for $\gamma_1, \ldots, \gamma_p$, then the PH assumption is reasonable.

16) not in text

Let $T_i = \min(Y_i, Z_i)$ and $\gamma_i = \begin{cases} 1 & T_i = Y_i \text{ uncensored} \\ 0 & T_i = Z_i \text{ censored} \end{cases}$

An ET plot is a plot of the ESP vs T_i where the plotting symbol is 0 if $\gamma_j = 0$ and + if $\gamma_j = 1$.

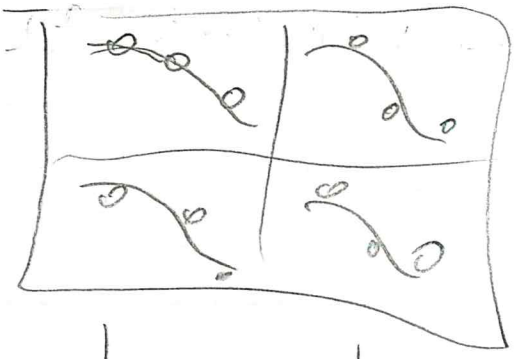
Since $h_i(t) = e^{ESP} h_0(t)$, hazard increases and survival decreases as ESP increases, if $ESP \simeq SP$.

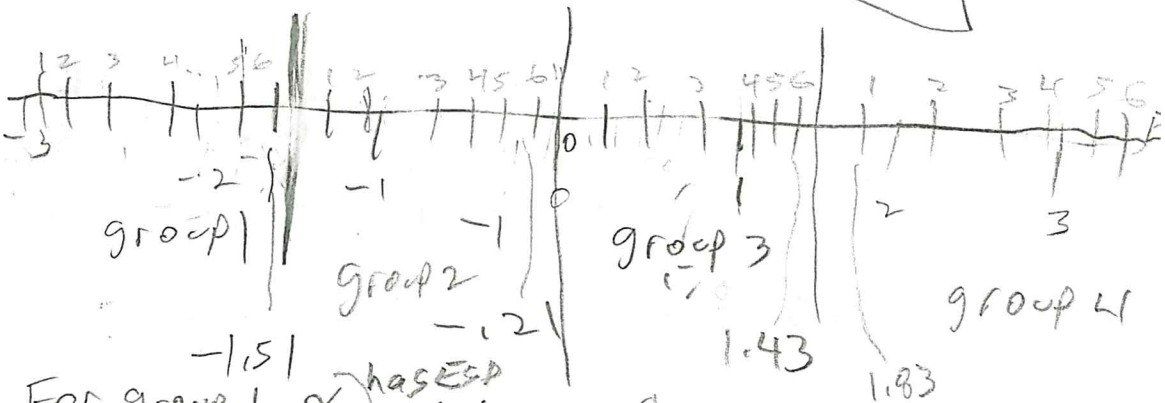


0 = censored
+ = death

Section 4ab

(17) know for final (not in text) The slice survival plot divides $ESP = \hat{\beta}'x$ into J groups with approx $\frac{n}{J}$ obs's. $\hat{S}_{PH_j}(t)$ is computed for each group using the x corresponding to the largest ESP. in the 1st $J-1$ groups and the x corresponding to the smallest ESP for the Jth group. The Kaplan Meier estimator $\hat{S}_{KM_j}$ is computed from the T_{ij} in each group. Then $\hat{S}_{PH_j}(t)$ is plotted and $\hat{S}_{KM_j}(t)$ is plotted at the jumps with circles as the plotting symbols. The PH assumption is reasonable if the circles track the curves well.





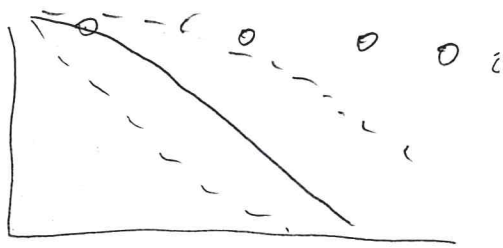
For group 1 x_{i1} has ESP $= -1.51$ and $\hat{S}_{KM_1}$ is computed from the 6 cases.

(18) If pointwise CI bands are added to $\hat{S}_{PH_j}$, $\hat{S}_{KM_j}$ tracks $\hat{S}_{PH_j}$ well if the most of the plotted circles do not fall very far outside the bands.

(It is possible to make confidence bands for the entire

curve) such that the plotted circles would fall entirely within the bands if the PH model was reasonable, but the "simultaneous bands" are much wider than the pointwise CI bands.)

good plot



pointwise CI bands ↑ bad plot

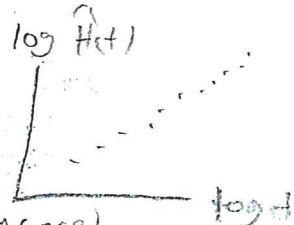
Ch8 1] pg54 If $h(t) = \lambda \gamma t^{\gamma-1}$ where $\lambda > 0, \gamma > 0$ and $t > 0$, then T has a Weibull (λ, γ) distribution. $S(t) = \exp(-\lambda t^{\gamma})$

2] pg57 $-\log S(t) = \lambda t^{\gamma}$ so

$$\log[-\log S(t)] = \log(\lambda) + \gamma \log(t).$$

3) pg57 If the data are from a censored Weibull dist, compute $\hat{S}_{KM}$. A plot of $\log(t)$ vs $\log(-\log \hat{S}_{KM}(t)) = \log(\hat{H}_{KM}(t))$ should be linear with slope γ and intercept $\log(\lambda)$.

$\gamma \approx 1 \Rightarrow$ data could be $EXP(\lambda)$.



§5.34) $Y_i^* = T_i = \min(Y_i, Z_i)$, $\delta_i = \begin{cases} 0 & \text{if } T_i = Z_i \text{ censored} \\ 1 & \text{if } T_i = Y_i \text{ uncensored} \end{cases}$

Suppose Z had pdf g_Z and CDF G_Z.

Let Y have pdf $f(y|\underline{\theta})$ and CDF $F(y|\underline{\theta})$ for $\underline{\theta} \in \Theta$.

So Y comes from a parametric family completely

determined by $\underline{\theta} \in (H)$.

ex] $f(y|\theta) = \lambda e^{-\lambda y}$, $y > 0$ $\quad \lambda \in (0,\infty) = (H)$.

5) P357-358 Want the MLE $\hat{\underline{\theta}}$ of $\underline{\theta}$ based on observed data. $(T_1,\delta_1), \dots, (T_n,\delta_n)$. The likelihood function $L(\underline{\theta}) = \prod_{i=1}^{n} f(t_i,\delta_i \mid \underline{\theta})$. Suppress $\underline{\theta}$.

T/S; right cont-nuous; discrete.

To find f, note that $F(t,0) = P(T \le t, \delta = 0) = P(Z \le t, Y > Z) = (*)$ ($T = \min(Y,Z) = Z$)

$P(T \le t, \delta \le 0) = P(T \le t, \delta = 0)$

Fact: $P(A) = E[P(A|Z)] = E[P(A|Y)]$.

$$S_0(*) = \int_0^t P(Z \le t, Y > Z \mid Z = v)\, g_Z(v)\, dv$$
$$= \int_0^t P(Z \le t, Y > v \mid Z = v)\, g_Z(v)\, dv$$
$$= \int_0^t \underbrace{P(Z \le t \mid Z = v)}_{1 \text{ since } v \le t \text{ on } (0,t)}\, P(Y > v)\, g_Z(v)\, dv. \quad S_0$$

$Y \perp\!\!\!\perp Z$

$F_{T/S}(t,0) = \int_0^t S_Y(v)\, g_Z(v)\, dv$. Differentiate both sides wrt t to get $f_{T/S}(t,0) = S_Y(t)\, g_Z(t)$ by the fundamental theorem of calculus. Now $\frac{d}{dt}(F(t,1) - F(t,0)) = f(t,1)$, and

($T = \min(Y,Z) = Y$) see notes 25

$F(t,1) - F(t,0) = P(T \le t, \delta = 1) = P(Y \le t, Y < Z)$

$$= \int_0^t P(Y \le t, Y < Z \mid Y = v)\, f_Y(v)\, dv$$
$$= \int_0^t P(Y \le t, v < Z \mid Y = v)\, f_Y(v)\, dv = \int_0^t \underbrace{P(Y \le t \mid Y = v)}_{\text{since } v \le t \text{ on } (0,t)}\, P(v < Z)\, f_Y(v)\, dv$$

$Y \perp\!\!\!\perp Z$

$$= \int_0^t (1 - G_Z(v))\, f_Y(v)\, dv.$$

Thus $f_{T/S}(t,1) = [1 - G_Z(t)]\, f_Y(t)$.

6) So $L(\underline{\theta}) = \prod_{i=1}^{n} f_{T,\delta}(t_i, \delta_i | \underline{\theta})$

$= f_Y(t_i|\underline{\theta})\, g_Z(t_i)$ if $\delta_i = 0$

$[1 - G_Z(t_i)]\, f_Y(t_i|\underline{\theta})$ if $\delta_i = 1$

$$= \prod_{i=1}^{n} \left[(1 - G_Z(t_i))\, f_Y(t_i|\underline{\theta}) \right]^{\delta_i} \left[S_Y(t_i|\underline{\theta})\, g_Z(t_i) \right]^{1-\delta_i}$$

$$= \underbrace{\prod_{i=1}^{n} [1 - G_Z(t_i)]^{\delta_i} [g_Z(t_i)]^{1-\delta_i}}_{\text{positive constant wrt } \underline{\theta}} \prod_{i=1}^{n} [f_Y(t_i|\underline{\theta})]^{\delta_i} [S_Y(t_i|\underline{\theta})]^{1-\delta_i}$$

$$\text{or } L(\underline{\theta}) = c \prod_{i=1}^{n} f_Y(t_i|\underline{\theta})^{\delta_i} [S_Y(t_i|\underline{\theta})]^{1-\delta_i}$$

Colleti p 159 takes $c=1$ since the value of $c>0$ will not affect the maximization, See p 358. consequence software log likelihoods $\log L$ can differ by a constant

=> p160 Censored Exp(λ) data. $f(t|\lambda) = \lambda e^{-\lambda t}$, $S(t|\lambda) = e^{-\lambda t}$

$$L(\lambda) = c \prod_{i=1}^{n} (\lambda e^{-\lambda t_i})^{\delta_i} (e^{-\lambda t_i})^{1-\delta_i} = c \prod_{i=1}^{n} \lambda^{\delta_i} e^{-\lambda t_i}$$

$$= c\, \lambda^{\sum \delta_i} e^{-\lambda \sum t_i}. \quad \text{So}$$

$$\log L(\lambda) = \log c + \sum \delta_i \log \lambda - \lambda \sum t_i$$

$$\frac{d \log L(\lambda)}{d\lambda} = \frac{\sum \delta_i}{\lambda} - \sum t_i \stackrel{\text{set}}{=} 0 \quad \text{or} \quad \lambda \sum t_i = \sum \delta_i$$

$$\text{or } \hat{\lambda} = \frac{\sum \delta_i}{\sum t_i}. \qquad \frac{d^2 \log L(\lambda)}{d\lambda^2} = -\frac{\sum \delta_i}{\lambda^2} < 0$$

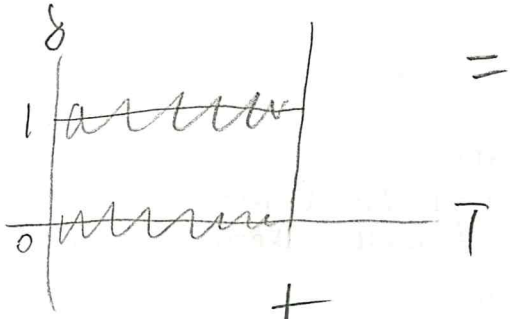
So $\hat{\lambda}$ is the MLE if $\sum t_i > 0$. By the invariance Principle,

$T \in (0, \infty)$ and $\delta \in \{0,1\}$

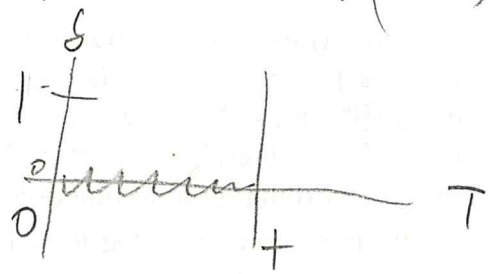
$F(t,1) = P(T \le t, \delta \le 1)$

$= F(t,1) = P(0 < T \le t, \delta = 1) + P(0 < T \le t, \delta = 0)$

$= P(T \le t, \delta = 1) + P(T \le t, \delta = 0)$



$F(t,0) = P(T \le t, \delta \le 0) = P(T \le t, \delta = 0)$



So $\frac{d}{dt} F(t,0) = f(t,0)$

and $\int_0^t f(u,0)\,du = F(t,0)$.

$F(t,1) - F(t,0) = P(T \le t, \delta = 1)$

So $\frac{d}{dt}\left[F(t,1) - F(t,0)\right] = f(t,1)$

and $\int_0^t f(u,1)\,du = F(t,1) - F(t,0)$.

$\sum_{j=0}^{1} \int_0^t f(u,j)\,du = F(t,0) + F(t,1) - F(t,0) = F(t,1)$

$= P(T \le t, \delta \le 1)$

↑ T has measure on $(0,\infty)$

↑ δ has "measure" on $\{0,1\}$

The derivative of the first term with respect to θ is 0.
ii) Find the derivative $\frac{d}{d\theta}\log L(\theta)$, set the derivative equal to zero and solve for θ. The solution is a candidate for the MLE.
iii) **Invariance Principle:** If $\hat{\theta}$ is the MLE of θ, then $\tau(\hat{\theta})$ is the MLE of $\tau(\theta)$.
iv) Show that $\hat{\theta}$ is the MLE by showing that $\hat{\theta}$ is the global maximizer of $\log L(\theta)$. Usually this is done by noting that $\hat{\theta}$ is the unique solution to the equation $\frac{d}{d\theta}\log L(\theta) = 0$ and that the 2nd derivative evaluated at $\hat{\theta}$ is negative: $\frac{d^2}{d\theta^2}\log L(\theta)|_{\hat{\theta}} < 0$.
See Q10, HW21 9.72ad, 9.73, 9.75, 9,76ad

Tips: a) $\exp(a) = e^a$ and $\log(y) = \ln(y) = \log_e(y)$ is the **natural logarithm.**
b) $\log(a^b) = blog(a)$ and $\log(e^b) = b$.
c) $\log(\prod_{i=1}^n a_i) = \sum_{i=1}^n \log(a_i)$.
d) $\log L(\theta) = \log(\prod_{i=1}^n g(y_i|\theta)) = \sum_{i=1}^n \log(g(y_i|\theta))$.
e) If t is a differentiable function and $t(\theta) \neq 0$, then $\frac{d}{d\theta}\log(|t(\theta)|) = \frac{t'(\theta)}{t(\theta)}$ where $t'(\theta) = \frac{d}{d\theta}t(\theta)$. In particular, $\frac{d}{d\theta}\log(\theta) = 1/\theta$.
f) Anything that does not depend on θ is treated as a constant with respect to θ and hence has derivative 0 with respect to θ.

Suppose $\frac{d}{d\theta}\log L(\theta_o) = 0$. The 2nd derivative test states that if $\frac{d^2}{d\theta^2}\log L(\theta_o) < 0$, then θ_o is a local max.

If $\log L(\theta)$ is strictly concave ($\frac{d^2}{d\theta^2}\log L(\theta) < 0$ for all θ), then any local max of $logL(\theta)$ is a global max.

You should know how to find the MLE for the normal distribution (including when μ or σ^2 is known, memorize the MLE's $\overline{Y}$, $\sum_{i=1}^n (Y_i - \overline{Y})^2/n$, $\sum_{i=1}^n (Y_i - \mu)^2/n$) and for the uniform distribution. Also $\overline{Y}$ is the MLE for several brand name distributions.

Confidence intervals: Confidence intervals are intervals of plausible values for the parameter and have the form estimator $\pm$ cutoff $\sqrt{V(estimator)}$ where V(estimator) is the estimated variance of the estimator. The cutoff $t_{\alpha/2}$ is obtained from the t-table (use the bottom of the t-table if the cutoff is $z_{\alpha/2}$).

TESTS OF HYPOTHESES WILL BE ON QUIZ 11 AND THE FINAL BUT NOT ON EXAM 4

tests of hypotheses: All tests of hypotheses have the same 4 steps
i) State Ho and Ha.
ii) Calculate the test statistic.
iii) Find the p–value.
iv) If the p–value $\leq \alpha$, reject Ho, otherwise fail to reject Ho. Write a nontechnical sentence explaining the decision.

Use the notes given in class for more details of the following procedures.

$T = \min(Y, Z)$ $\quad \delta = 1$ if $Y \le Z$ uncensored

$\delta = 0$ if $Y > Z$ censored

Z had pdf g_Z DF G_Z

(compare Miller p17 Also Cohett p357 -358)

$$P(T \le t, \delta = 0) = P(Z \le t, Y > Z)$$

$$= \int_0^t P(Z \le t, Y > Z \mid Z = u)\, g_Z(u)\, du$$

↑ q

goes with notes 28

$P(A) = E(I_A) = E[E(I_A \mid Z)] = E(P(A \mid Z))$

$$= \int_0^t P(Z \le t, Y > u \mid Z = u)\, g_Z(u)\, du$$

$$= \int_0^t \underbrace{P(Z \le t \mid Z = u)}_{1 \text{ since } u \le t \text{ on } (0,t)}\, P(Y > u)\, g_Z(u)\, du$$

↑ $Y \perp\!\!\!\perp Z$

$$= \int_0^t S_Y(u)\, g_Z(u)\, du$$ So differentiate both sides

and $f(t,0) = S_Y(t)\, g_Z(t)$ by the Fund Th of Calc.

Similarly $P(T \le t, \delta = 1) = P(Y \le t, Y < Z)$

$$= \int_0^t P(Y \le t, Y < Z \mid Y = u)\, f_Y(u)\, du$$

$$= \int_0^t P(Y \le t, u < Z \mid Y = u)\, f_Y(u)\, du$$

$Y \perp\!\!\!\perp Z$

$$= \int_0^t \underbrace{P(Y \le t \mid Y = u)}_{1 \text{ since } u \le t \text{ on } (0,t)}\, P(u < Z)\, f_Y(u)\, du$$

$$= \int_0^t [1 - G_Z(u)]\, f_Y(u)\, du$$

So $f(t,1) = [1 - G_Z(t)]\, f_Y(t)$.

(Smith uses G as a survival function, not as a CDF.)

ignore

Math 483 EXAM 1 covers 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 2.8, 2.9, 3.1, 3.2, 3.3, and 3.4 and is on WEDNESDAY, FEB. 9. You are allowed TWO SHEETS OF NOTES and a CALCULATOR.

A **set** consists of distinct elements enclosed by *braces*, eg $\{1, 5, 7\}$.

The *universal set* S is the set of all elements under consideration.

The *empty set* $\emptyset$ is the set that contains no elements.

A is a subset of B, $A \subseteq B$, if every element in A is in B.

The **union** of A with $B = A \cup B$ is the set of all elements in A or B (or in both).

The **intersection** of A with $B = A \cap B$ is the set of all elements in A and B.

If $A \cap B = \emptyset$, then A and B are (**mutually exclusive** or) **disjoint sets.**

The *complement* of A is $\overline{A}$, the set of elements in S but not in A.

Know DeMorgan's Laws.

The *sample space* S is the set of all possible outcomes of an experiment. A *sample point* E_i is a possible outcome. An *event* is a subset of S. A simple event is a set that contains exactly one element of S, eg $A = \{E_3\}$. A *discrete sample space* consists of a finite or countable number of outcomes.

The *relative frequency interpretation of probability* says that the probability of outcome (sample point) E_i is the proportion of times that E_i would occur if the experiment was repeated again and again infinitely often.

For **any event** A, $0 \leq P(A) \leq 1$.

Three axioms: $P(A) \geq 0, P(S) = 1$, and if $A_1, A_2, ...$ are pairwise mutually exclusive, then $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

1) **Common problem.** Use order to find S. Using a table to find S if two die are tossed or if a die is tossed twice and to find S if a coin is flipped 2, 3, or 4 times are typical examples. See Q1 6, HW2 extra problem.

The *sample point method* for finding the probability for event A says that if $S = \{E_1, ..., E_k\}$ then $0 \leq P(E_i) \leq 1$, $\sum_{i=1}^{k} P(E_i) = 1$, and $P(A) = \sum_{i:E_i \in A} P(E_i)$. That is, $P(A)$ is the sum of the probabilities of the sample points in A. If all of the outcomes E_i are *equally likely,* then $P(E_i) = 1/k$ and $P(A) =$ (number of outcomes in A)$/k$ if S contains k outcomes.

2) **Common Problem.** Leave the probabilities of some outcomes blank. See Q1 1 and HW2 2.13.

3) **Common problem.** List all outcomes in S and use these outcomes to find $P(A)$. See HW2 extra problem.

The *multiplication mn rule* says that if there are n_1 ways to do a first task, n_2 ways to do a 2nd task, ..., and n_k ways to do a kth task, then the number of ways to perform the total act of performing the 1st task, then the second task, ..., then the kth task is $n_1 \cdot n_2 \cdot n_3 \cdots n_k$. Techniques for multiplication principle: a) use a slot for each task

The MLE of $E Y_i = \frac{1}{\lambda}$ is $\frac{1}{\hat{\lambda}} = \frac{\sum_{i=1}^{n} t_i}{\sum_{i=1}^{n} \delta_i}$.

Let $r = \sum_{i=1}^{n} \delta_i$ = # uncensored cases. Then

$$\frac{1}{\hat{\lambda}} = \frac{\sum_{i=1}^{n} t_i}{r} = \frac{\text{total observed time survived}}{\text{\# deaths observed}}$$

7) Know for E3 p161 See HW8.1 For EXP(λ) data, $SE\left(\hat{\lambda}\right) = \frac{\hat{\lambda}}{\sqrt{r}}$

So a 95% CI for λ is $\hat{\lambda} \pm 1.96\, SE(\hat{\lambda})$

$$= \hat{\lambda} \pm 1.96 \frac{\hat{\lambda}}{\sqrt{r}}.$$

A better CI is $\left(\frac{\hat{\lambda}\, \chi^2_{2r,1-\alpha/2}}{2r}, \frac{\hat{\lambda}\, \chi^2_{2r,\alpha/2}}{2r} \right)$

ex] Suppose the censored data are from an EXP(λ) dist. a) Find $\hat{\lambda}$ and b) a 95% CI for λ.

Miller p186 1, 1, 1, 1+, 4+, 5, 7, 8, 10, 10+, 12+, 16+, 16+, 16+

(# uncensored = r = n − # censored; n → 14, # censored → 7)

a) $\hat{\lambda} = \frac{\sum \delta_i}{\sum t_i} = \frac{\sum_{i=1}^{n} \delta_i}{\sum_{i=1}^{n} t_i} = \frac{7}{1+1+1+1+4+5+7+8+10+10+12+16+16+16}$

$$= \frac{7}{108} \approx 0.06481$$

(7 ← # uncensored)

b) $r = 7$ $\hat{\lambda} \pm 1.96 \frac{\hat{\lambda}}{\sqrt{r}} = 0.06481 \pm 1.96\left(\frac{0.06481}{\sqrt{7}}\right)$

$= 0.06481 \pm .048015 = (0.01679, 0.1128)$

Skim §5.4, it is a special case of §5.5

8) * p176 A <u>parametric PH model</u> has the form

$h_{\underset{\sim}{x}}(t) = \exp(\underset{\sim}{\beta}_w' \underset{\sim}{x})\, h_0(t|\underset{\sim}{\theta})$. Here $h_0(t|\theta)$ is completely known given θ. (For the semiparametric Cox PH model, $h_0(t)$ is left unspecified.)

9) p176 The Weibull PH model uses $h_0(t) = \lambda\gamma t^{\gamma-1}$, so $\theta = (\lambda, \gamma)$. The exponential PH model is the special case with $\gamma = 1$, so $h_0(t) = \lambda$. Here $\underset{\sim}{\theta} = \theta = \lambda$.

10) p176 p289, 273 For the Weibull PH model, $Y|SP \sim W(\lambda e^{SP}, \gamma)$.

and $S_{\underset{\sim}{x}}(t) = \exp\left[-\exp(SP)\lambda t^{\gamma}\right]$ where $SP = \underset{\sim}{\beta}_w' \underset{\sim}{x}$.

So $H_{\underset{\sim}{x}}(t) = \exp(SP)\lambda t^{\gamma} = \int_0^t h_{\underset{\sim}{x}}(u)\,du$.

11) * p178 p275 Instead of fitting a WPH model, SAS and R fit an accelerated failure time log linear model

$\log Y_i = \alpha + \underset{\sim}{\beta}' \underset{\sim}{x}_i + \sigma\varepsilon_i$ where $\text{Var}\,\varepsilon_i = 1$ ($E\varepsilon_i = 0$ since α is in the model)

Here $\underset{\sim}{\beta} \neq \underset{\sim}{\beta}_w$, but $\underset{\sim}{\beta}_w = -\underset{\sim}{\beta}/\sigma$.

Basically if $Y \sim W(\lambda, \gamma)$ then $\log Y$ has a smallest extreme value distribution and the log likelihoods of Y and $\log Y$ differ by a constant.

12) $\hat{\lambda}$ and $\hat{\underset{\sim}{\beta}}$ are MLEs Have $(T_i, \delta_i, \underset{\sim}{x}_i)$ not $(Y_i, \underset{\sim}{x}_i)$.

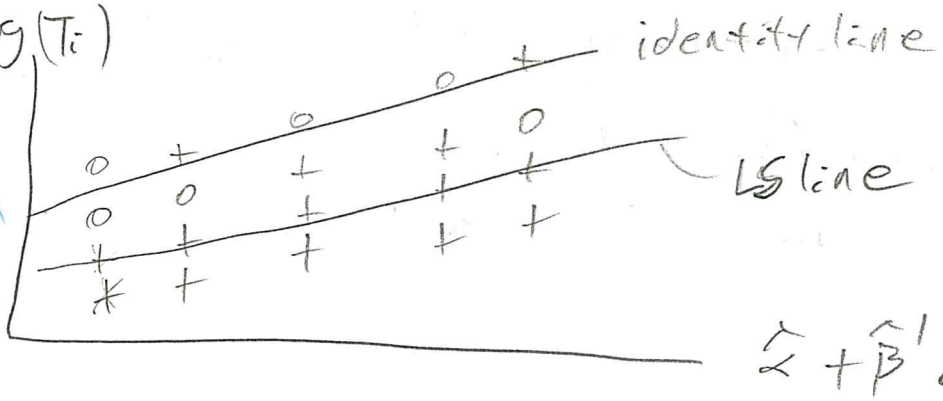
Let $\log T_i = \hat{\lambda} + \hat{\underset{\sim}{\beta}}' \underset{\sim}{x}_i + r_i$

(13) *Plot of $\hat{\alpha} + \hat{\beta}' x_i$ vs $\log T_i$ with plotting symbol

0 for $\delta_i = 0$ and + for $\delta_i = 1$.

If the weibull PH model is reasonable, the plotted points will scatter about the identity line with vertical deviation $= \log(T_i) - (\hat{\alpha} + \hat{\beta}' x_i) = r_i$

Also add the least squares line based on the uncensored cases to the plot. Want the LS line to have a slope not too far from 1. If $\hat{\beta} \approx \beta$, then want $\hat{\alpha} + r_i \approx \alpha + \sigma \varepsilon_i$. Typically the ε_i are left skewed so the identity line is higher than most of the plotted cases. Since $V(\alpha + \sigma \varepsilon_i) = \sigma^2$, want the r_i to have roughly constant variance



big problem

↓

(Long left tail of SEV dist makes outliers and influential cases hard to detect.)

(14) know for final Given output and x_j, be able to compute

$ESP = \hat{\beta}' x = \hat{\beta}_1 x_1 + \cdots \hat{\beta}_p x_p$. given x.

(15) know for final p277 A large sample 95% CI for β_j is $\hat{\beta}_j \pm 1.96 SE(\hat{\beta}_j)$

(16) know for final p293 the 4 step wald test of hypotheses is the same i) $H_0: \beta_j = 0$ H_A $\beta_j \neq 0$

ii) $z_{0j} = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)}$ or $X^2_{0j} = Z^2_{0j}$ often from output

iii) $pval = 2P(Z < -|Z_{0j}|) = P(\chi_1^2 > \chi_{0j}^2)$

often from output.

iv) If $pval < \delta$ reject H_0 and conclude X_j is needed in the Weibull survival model given the other $p-1$ predictors are in the model. If $pval \geq \delta$ fail to reject H_0 and conclude X_j is not needed in the Weibull survival model given the other $p-1$ variables are in the model.

17) P279 Know for Final The 4 step likelihood ratio test LRT is

i) H_0 $\underline{\beta} = \underline{0}$ H_A $\underline{\beta} \neq 0$

ii) $\chi^2(MF) = [-2\log L(\text{none})] - [-2\log L(\text{full})]$

(This is Chisq from R. Both R and SAS give $\log L$ [but actually $\log L_R = \log L + d_R$ and $\log L_{SAS} = \log L + d_{SAS}$].)

iii) $pval = P(\chi^2_p > \chi^2(MF))$. Note that $p+2-2 = p$.

iv) Reject H_0 if $pval < \delta$ and conclude there is a WPH survival relationship between Y and the predictors $\underline{X}$. Fail to reject H_0 if $pval \geq \delta$ and conclude there is not a WPH relationship between Y and the predictors $\underline{X}$.

18) Note that there could be a PH survival relationship but not a WPH survival relationship. Check the WPH assumptions before doing inference.

19) know for final pick one 4 step change in

LR test is i) H_0 the reduced model is good H_A use the full mo

ii) $\chi^2(RIF) = \chi^2(WIF) - \chi^2(NIR) = [-2\log L(red)] - [-2\log L$

(Get $\chi^2(NIF)$ and $\chi^2(NIR)$ from Chisq in R or get $\log L(red)$ and $\log L(full)$ from R or SAS):

iii) $pval = P(\chi^2_{p-r} > \chi^2(RIF))$ where

the full model has $p-r$ more predictors than the reduced model. Note that $p-r = (p+2) - (r+2)$.

iv) reject H_0 if $pval < \delta$ and conclude the full model should be used. Fail to reject H_0 if $pval \geq \delta$ and conclude the reduced model is good.

ex) ovarian cancer ~~p187-190~~ and ~~344-346~~
see output on ex3 review p2-3 (full model)

a) Find ESP if treat=1 and age = 60.

Soln) $ESP = \hat{\beta}'x = -0.561(1) - 0.079(60) = -5.301$
R output

b) Find a 95% CI for β_1 corresponding to treat

Soln) $\hat{\beta}_1 \pm 1.96\, SE(\hat{\beta}_1) = -0.561 \pm 1.96(0.3399) = -.561 \pm .666$

$(-1.2272, .1052)$

c) Test $\beta_1 = 0$ corresponding to treat

i) $H_0\ B_1 = 0$ $\quad H_A\ B_1 \neq 0$

ii) $z_{01} = \dfrac{\hat{B}_1}{SE(\hat{B}_1)} = \dfrac{-.561}{.3399} = -1.6504$ (or use output

or $X^2_{01} = z^2_{01} = 2.7241$ (2.73 from output)

Tiii) $pval = 2P(Z \leq -1.65) = 2(.0495) = .099$
(.0986 from output)

or $pval = P(\chi^2_1 > 2.72)$

df	.100	.05
1	2.71	3.841

so $.05 \leq pval < .100$ (.0986 from output)

iv) Fail to reject H_0 treatment is not needed in the Weibull survival model given age is in the model.

d) Test $B_2 = 0$ corresponding to age.

i) $H_0\ B_2 = 0$ $\quad H_A\ B_2 \neq 0$

ii) $z_{02} = -4.00$
or $x^2_0 = 15.97$

iii) $pval = .0000643$
or $pval < .0001$

iv) reject H_0 age is needed in the weibul survival model given treat is in the model.

e) Test $\underset{\sim}{B} = \underset{\sim}{0}$

ii) Ho [illegible] HA [illegible]

ii) R: $X^2(WIF) = 18.41$

or $X^2(WIF) = [-2\log L(none)] - [-2\log L(full)]$

$= [-2(-98)] - [-2(-88.7)] = 196 - 177.4$

$= 18.6$ (rounding)

or SAS $X^2(WIF) = [-2(-29.7672)] - [-2(-20.563$

$= 59.5344 - 41.1262 = 18.4082$

iii) $pval = P(\chi^2_2 > 18.41)$

df	.001
2	13.82

$pval < .001$ (.0001 from oct pct)

iv) reject Ho there is a WPH survival relationship between time Y and the predictors age and treat.

f) Test whether reduced model with treat is good

i) Ho the reduced model is good HA use the full model

ii) $X^2(RIF) = X^2(WIF) - X^2(WIR) = 18.41 - 1.18 = 17.23$

R

or $X^2(RIF) = [-2\log L(Red)] - [-2\log L full] =$

R: $X^2(RIF) = [-2(-97.4)] - [-2(-88.7)] = 194.8 - 177.4 = 17.4$ (rounding)

SAS: $X^2(RIF) = [-2(-29.1775)] - [-2(-20.5631)] = 58.355 - 41.126$

$= 17.2288$

iii) $pval = P(\chi^2_1 > 17.23)$ $pval < .001$

df	.001
1	10.83

iv) reject Ho, use the full model

25) SAS and R do not have direct methods for variable selection, but the WPH model is a PH model. So put the data in SAS, use PH variable selection to select models. Fit the Weibull regression submodels and check WPH assumptions for the candidate submodels.

1] §6.3 is a special case of §6.4 since skim §6.3

§6.4 2] I know for final ~~p207~~ p272-5 The accelerated failure time AFT model is

$$\log Y_i = \alpha + \underline{\beta}'\underline{x}_i + \sigma\varepsilon_i = \alpha + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \sigma\varepsilon_i$$

Assume $x_i > 0$ and Note that $\beta_i > 0$ increases $\log Y_i$ and Y_i while $\beta_i < 0$ decreases $\log Y_i$ and time Y_i.

(Collett uses $\log T_i = \mu + \alpha_1 x_{i1} + \cdots + \alpha_p x_{pi} + \sigma\varepsilon_i$.)

For Cox PH model $\lambda_{\underline{x}}(t) = e^{\underline{\beta}'\underline{x}} h_0(t)$ so

$\beta > 0$ increases hazard and decreases Y_i while

$\beta < 0$ decreases hazard and increases Y_i.

So in WPH model $\underset{\sim}{\beta}_W = -\underset{\sim}{\beta}/\sigma$.

3] The Weibull and Exponential regression models are the only models that are both PH and AFT models. (for any log location scale family, the lognormal and loglogistic

4) p206 $h_{\underline{x}}(t) = h_i(t) \equiv h_{\underline{x}|sp}(t) = e^{-sp}\, h_0(e^{-sp}t)$

$= \frac{1}{e^{sp}}\, h_0\left(\frac{t}{e^{sp}}\right)$ for the AFT model.

5) $S_{\underline{x}|sp}(t) \equiv S_i(t) \equiv S_{\underline{x}}(t) = S_0\left(\frac{t}{e^{sp}}\right) = S_0(e^{-sp}t)$

6) p209 Let $F_0(t_{0,\rho}) = P(Y \le t_{0,\rho} \mid sp = \underline{\beta}'\underline{0} = 0) = \rho$

So $S_0(t_{0,\rho}) = 1 - F_0(t_{0,\rho}) = 1-\rho$.

Let $F_{\underline{x}}(t_{\underline{x},\rho}) = P(Y \le t_{\underline{x},\rho} \mid sp = \underline{\beta}'\underline{x}) = \rho$

So $S_{\underline{x}}(t_{\underline{x},\rho}) = 1-\rho = S_0\left(\frac{t_{\underline{x},\rho}}{e^{sp}}\right) = S_0(t_{0,\rho})$

So $t_{0,\rho} = \frac{t_{\underline{x},\rho}}{e^{sp}}$ or $t_{\underline{x},\rho} = e^{sp}\, t_{0,\rho}$

7) p207 $\log Y_i = \underbrace{\alpha}_{\text{intercept parameter}} + \underline{\beta}'\underline{x} + \underbrace{\sigma}_{\text{scale parameter}}\varepsilon_i$

$S_{\underline{x}}(t) = \underset{sp=\underline{\beta}'\underline{x}_i}{P}(Y_i > t) = P\left[\exp(\alpha + \underline{\beta}'\underline{x}_i + \sigma\varepsilon_i) > t\right]$

$= P\left(\exp(\alpha+\sigma\varepsilon_i)\exp(sp) > t\right) = P\left(\exp(\alpha+\sigma\varepsilon_i) > \frac{t}{e^{sp}}\right)$

$= P_{sp=\underline{\beta}'\underline{0}}\left(Y_i > \frac{t}{e^{sp}}\right) = S_0\left(\frac{t}{e^{sp}}\right)$.

8) p208 $H_{\underline{x}}(t) = -\log S_{\underline{x}}(t) = -\log S_0\left(\frac{t}{e^{sp}}\right) = H_0\left(\frac{t}{e^{sp}}\right)$. So $h_{\underline{x}}(t) = \frac{d}{dt} H_{\underline{x}}$

$= \frac{d}{dt} -\log S_0\left(\frac{t}{e^{sp}}\right)$ $= \frac{d}{dt} H_0\left(\frac{t}{e^{sp}}\right)$

$= h_0\left(\frac{t}{e^{sp}}\right)\frac{1}{e^{sp}}$.

9) From 6) median survival time

$t_{\underline{x},.5} = e^{sp}\, t_{0,.5}$. If $e^{sp} > 1$, the median survival time of $\underline{x}$ > median survival time of $\underline{0}$. This is good if event = death, bad if event = time until recovery, The acceleration factor $= e^{-sp}$ and $t_{\underline{0},p} = e^{-sp}\, t_{\underline{x},p}$, $S_{\underline{x}}(t) = S_0(e^{-sp}t)$.

Here $S_{\underline{x}}(t_{\underline{0},p}) = 1-p = P(Y > t \mid SP = \underline{\beta}'\underline{x})$, $0 < p < 1$.

10) For the Exponential AFT regression model, $\sigma = 1$ and ε has a standard SEV(0,1) distribution. For the Weibull AFT regression model σ is unknown but $\varepsilon \sim$ SEV(0,1),

Here $f_\varepsilon(t) = e^t e^{-e^t} = \exp(t - e^t)$ for $t \in (-\infty, \infty)$.

($e^a e^b = e^{a+b}$; t = log(time))

The SEV(0,1) dist has a long left tail

$E(\varepsilon) \approx -0.5772$ $V(\varepsilon) = \frac{\pi^2}{6} \approx 1.6449$.

KEY WORDS: Censored Data; Median Absolute Deviation; Method of Moments.

11) ~~p 210~~ For the Weibull AFT, $Y \sim W(\lambda e^{-\delta sp}, \gamma)$

For the Exponential AFT $Y \sim Exp(\lambda e^{-sp}) \stackrel{D}{=} W(\lambda e^{-sp}, 1)$.

12) For the lognormal AFT

§8.4

$$\log Y_i = \alpha + \underline{\beta}'\underline{x} + \sigma\varepsilon_i \sim N(\alpha + \underline{\beta}'\underline{x}, \sigma^2)$$

So $Y_i \sim LN(\alpha + \underline{\beta}'\underline{x}, \sigma^2)$. $Var(\log Y_i) = \sigma^2$.

13) p212 For the loglogistic AFT

§8.4

$$\log Y_i = \alpha + \underline{\beta}'\underline{x} + \sigma\varepsilon_i \sim \text{logistic}(\alpha + \underline{\beta}'\underline{x}, \sigma)$$

where $\varepsilon_i \sim \text{logistic}(0,1)$ with $E\varepsilon_i = 0$ $\text{var}\varepsilon_i = \frac{\pi^2}{3}$.

$$Var(\log Y_i) = \frac{\sigma^2\pi^2}{3}.$$

$Y_i \sim \text{loglogistic}(\theta - \kappa sp, \kappa)$, $\kappa = \frac{1}{\sigma}$, $\theta = -\alpha/\sigma$.

14) The lognormal and loglogistic AFT models are not PH models (and Y does not follow a PH model).

15) Know for final ~~p216~~ For the AFT perform inference as done for WPH (equivalent Weibull AFT). So ESP, CI for β_i, Wald test for β_i, LR test for $H_0: \underline{\beta} = \underline{0}$ and change in LR test for H_0 reduced model is good are the same. Output for AFT models will include an intercept ($\hat{\alpha}$) and estimate of scale ($\hat{\sigma}$).

16) p 223, 225 (p301) The log logistic model is also a proportional odds (PO) model

$$\frac{S_{\underline{x}}(t)}{1-S_{\underline{x}}(t)} = e^{SP}\,\frac{S_{\underline{0}}(t)}{1-S_{\underline{0}}(t)} \quad \text{where}$$

$SP = \underline{\beta}_{PO}'\underline{x}$. $\quad \underline{\beta}_{PO} = \dfrac{\underline{\beta}_{AFT}}{\sigma}$.

The log logistic model is the only AFT model that is also a PO model.

1) Note that $\dfrac{S_{\underline{x}}(t)}{1-S_{\underline{x}}(t)} = \dfrac{P(Y>t\mid \underline{x})}{1-P(Y>t\mid \underline{x})}$ = odds of survival beyond time t

The log odds ratio $\log\left[\dfrac{\left(\frac{S_{\underline{x}}(t)}{1-S_{\underline{x}}(t)}\right)}{\left(\frac{S_{\underline{0}}(t)}{1-S_{\underline{0}}(t)}\right)}\right] = \underline{\beta}_{PO}'\underline{x}$.

Ch6 Checking reg models

Ch7 } p100, 120) For Cox PH $\hat{S}_{\underline{x}}(t) = \left[\hat{S}_{\underline{0}}(t)\right]^{e^{\hat{\beta}_C'\underline{x}}}$

where $\hat{S}_0$ and $\hat{\beta}_C$ are semiparametric estimators.

For parametric PH $\hat{S}_{\underline{x}}(t) = \left[\hat{S}_{\underline{0}}(t)\right]^{e^{\hat{\beta}_P'\underline{x}}}$ where $\hat{S}_{\underline{0}}$ and $\hat{\beta}_P$ are parametric estimators.

For AFT, $\hat{S}_{\underline{x}}(t) = \hat{S}_{\underline{0}}\left(\dfrac{t}{e^{\hat{\beta}_A\underline{x}}}\right)$.

2) Not in text The slice survival plot

ii). 6) For each of the three pairs of plots below, determine which model is better, i) or

divides $\hat{\beta}'\underline{x}$ into J groups, and plots $\hat{S}_{\underline{x}_j}(t)$ for $j=1,\ldots,J$ where $\underline{x}_j$ is from the jth group. Circles

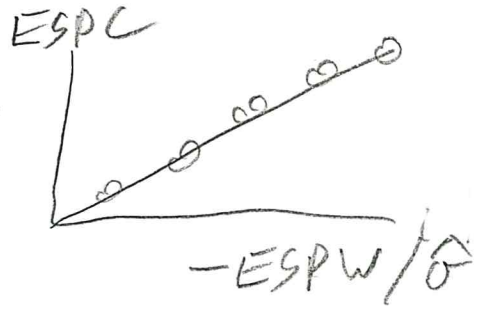
to the plot where $\hat{S}_{KMj}$ is the KM estima computed from the cases in the jth group. The survival model gives a reasonable fit if the circles $\hat{S}_{KMj}$ track the survival curves $\hat{S}_{x_j}(t)$ well.

3) Not in text. For the Weibull AFT, β_{WPH} $\underline{\beta}_C = -\underline{\beta}_A/\sigma$. For the Exponential AFT β_{EPH} $\underline{\beta}_C = -\underline{\beta}_A$. The Weibull EE plot is a plot of $\frac{-1}{\hat{\sigma}} ESPW = \frac{-1}{\hat{\sigma}} \hat{\underline{\beta}}_A' \underline{x} = \hat{\underline{\beta}}_{WPH}' \underline{x}$ vs

$ESPC = \hat{\underline{\beta}}_C' \underline{x}$. The Exponential EE plot is a plot of $-ESPE = -\hat{\underline{\beta}}_E' \underline{x}$ vs $ESPC = \hat{\underline{\beta}}_C' \underline{x}$ $= \hat{\underline{\beta}}_{EPH}' \underline{x}$

If the parametric model is good, as $n \to \infty$ the plotted points track the identity line with correlation $\to 1$.

ex)



Weibull reg is good but Exp reg is bad.

4) $\log Y_i = \alpha + \underline{\beta}_A' \underline{x}_i + \sigma \varepsilon_i$

$\log T_i = \hat{\alpha} + \hat{\underline{\beta}}_A' \underline{x}_i + r_i$

$\frac{r_i}{\hat{\sigma}}$ is a standardized residual

5) For the Exponential regression model, let $U = \sum_{i=1}^{n} \gamma_i$ = # uncensored cases.

Let $\hat{T}_i = \exp(\hat{\alpha} + \hat{\underline{\beta}}' \underline{x}_i)$. Then $\hat{\alpha}$ is such that

$U = \sum_{i=1}^{n} \frac{T_i}{\hat{T}_i}$. If $\frac{T_i}{\hat{T}_i} \approx c$, then

$nc = U$ or $c = \frac{U}{n}$ or $\frac{T_i}{\hat{T}_i} \approx \frac{U}{n}$ or $\hat{T}_i \approx \underbrace{\frac{n}{U}}_{\geq 1} T_i$

roughly. So in a plot of $\hat{\alpha} + \hat{\beta}' x_i$ vs $\log T_i$, the identity line may be higher than most of the plotted points

This may imply $\hat{d} = \hat{\alpha} - \alpha \xrightarrow{D} d \neq 0$,

but in simulations, $\hat{\alpha}$ seemed to be a consistent estimator of α.

6) Programming $\hat{S}_{\underline{x}}(t)$ so the parametric Slice Survival Plots look good for parametric models may be hard. So the EE plot is better for Exponential and Weibull regression. Use the cox slice survival plot to check PH assumptions, then use the EE plot to check the parametric distribution.

log cumulative hazard plot of $\log r_{ci}$ vs $-\log \hat{S}_{KM}(r_{ci})$ should follow the identity line if the fitted survival model is correct.

(In the HW, you will compare Fig 7.1 to the Weibull EE plot.)

8) §7.3 p 240-241 is basically the slice survival plot but all J curves are put on one plot without pointwise CI bands. Also $\hat{S}_j = \frac{1}{n_j}\sum_{i \in \text{Slice}_j} \hat{S}_{x_i}(t)$ = ave of estimated survival curves in each group.

9) Simulation generates pseudo survival regression data. Use the data to check programs and to find out how large $\frac{n}{p}$ should be for plots to be effective.

10) Let $\lambda = e^{\beta' x}$ and $V \sim Exp(\lambda)$. Then $EV = \frac{1}{\lambda}$.

Let $Y = g(V)$ where $g(v) > 0$ and $g^{-1}(y) > 0$ exists

$P(V > t) = e^{-\lambda t}$.

$$P(Y > t) = P(g(V) > t) = P(V > g^{-1}(t)) = e^{-\lambda g^{-1}(t)} = S_Y(t)$$

Since $S_Y(t) = \exp(-H_Y(t))$, $H_Y(t) = \lambda g^{-1}(t) = e^{\beta' x} g^{-1}(t)$

If $Y = V^{1/k} = g(V)$, then $g^{-1}(y) = y^k$!

Then $H_Y(t) = e^{\beta' x} t^k$ and $h_Y(t) = \frac{d}{dt} H_Y(t) = e^{\beta' x} k t^{k-1}$.

$X \sim W(\lambda, \gamma) \Rightarrow h_X(t) = \lambda \gamma t^{\gamma - 1}$, so $Y | x \sim W(\lambda = e^{\beta' x}, \gamma = k)$

To get censored data, let $T_i = \min(Y_i, Z_i)$ where $Z_i \sim Exp(1) \perp Y_i$

To get $SP = \underline{\beta}_P' \underline{x}_i$ let x_{ij} be pseudo $N(0,1)$

and $\underline{\beta}_P = (1, \ldots, 1)'$, $Y | SP \sim W(\lambda = e^{SP}, \gamma = k)$

is a weibull PH model. $\gamma = k = 1$ gives

a Exponential PH model.

11] For the Weibull AFT, $\gamma = \frac{1}{\sigma}$ and $\lambda = \exp\left(\frac{-\alpha}{\sigma}\right)$.

For the simulated data in 10], $\alpha = 0$ and

$\gamma = k = \frac{1}{\sigma}$ so $\sigma = \frac{1}{k}$.

$\underline{\beta}_P = -\frac{\underset{\sim}{\beta}_A}{\sigma}$ so $\underline{\beta}_A = -\sigma \underline{\beta}_P = -\left(\frac{1}{k}, \ldots, \frac{1}{k}\right)' = -\frac{1}{k} \underset{\sim}{1}$.

For data from 10]

12] and Cox regression, want $\hat{\underset{\sim}{\beta}}_C \approx \underline{\beta}_C = \underline{\beta}_P = (1, \ldots, 1)'$.

For weibull AFT want $\hat{\alpha} \approx 0$, $\hat{\underset{\sim}{\beta}}_A \approx -\frac{1}{k}(1, \ldots, 1)'$, $\hat{\sigma} \approx \frac{1}{k}$. See Hw9 wregsim3.

13) Collett p210-211 For data from 10],

$$\hat{S}_{Y|\underline{x}}(t) = \exp\left[-\exp(\hat{\underline{\beta}}_P' \underline{x})\, \hat{\lambda}\, t^{\hat{\gamma}}\right]$$

$$= \exp\left[-\exp\left(\frac{-\hat{\underline{\beta}}_A' \underline{x}}{\hat{\sigma}}\right) \exp\left(\frac{\log t - \hat{\alpha}}{\hat{\sigma}}\right)\right].$$

14) If the phdata program does too much censoring, there are numerical problems.

15 sim $ 7.4

§7.5 15) The Weibull model is the most commonly used parametric PH model, but the Cox semiparam PH model is used much more often for survival analys

16) Put Weibull regression data into cox PH software and use cox.zph (plot of scaled Schoenfeld residuals vs x_i with loess added as a visual aid, loess curvesshould be horizontal if PH assumption is reasonable) and Slice Survival plot to check PH assumpt. If PH assumption is reasonable, make the EE plot to check Weibull assumption.

ch 11 §11.1 1) Some times the Cox PH model does not fit the data as a whole, but there is a categorical variable A such that a Cox PH model fits each group = level of A. (not a stratified PH model)

ex) A = gender $h_{x,F}(t) = e^{\beta_F' x} h_{0F}(t)$

and $h_{x,M}(t) = e^{\beta_M' x} h_{0M}(t)$.

2) * p 215 The Stratified proportional hazards model is not a proportional hazards model but has $h_{x,j} = e^{\beta' x} h_{0j}(t)$ for $j = 1, \ldots, J$

for the J groups. Notice that the same $\underline{\beta}$ is used for each group (stratum) but each group has a different baseline function $h_{0j}(t)$. The model in point 1) had a different $\underline{\beta}_j$ for each group.

3) For each group j, a PH model $h_{\underline{x}}(t) = e^{\underline{\beta}'\underline{x}} h_{0j}(t)$ holds.

4) p316 know for final Inference is done exactly the same as for the Cox PH model. The output looks exactly the same, too. (You need to be told that the output is from a Cox PH model or a stratified PH model). But the model is a stratified PH model

so to test $\underline{\beta} = \underline{0}$ and $\beta_i = 0$ change PH to SPH in the conclusion.

ex) The dataset in Quiz 4, HW5.1 and HW7.1 was actually a stratified PH model data set. This dataset was lung cancer data and was stratified by sex. So the 2 groups = strata were F and M.

5) ~~For~~ the stratified PH model, NO $\hat{\beta}$'s are produced for the J levels of the

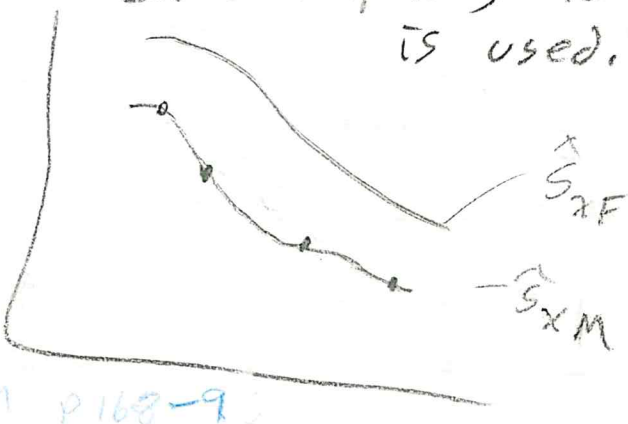
variable that defines

the strata. A Cox regression would use J-1 indicator variables. for the categorical variable.

6) To get information on the strata variable, make survival plots corresponding to each level for various $\underline{x}$.

Often $\underline{x}_a$ corresponding to the average of the $\underline{x}_i$ $i=1,\dots,n$ is used.

ex}



Allison p168-9

Better way! divide $\hat{Exp} = \hat{\underline{\beta}}'\underline{x}$ into 4 or 9 groups, make a plot for an $\underline{x}_i$ from each group.

7) Application: Suppose there are 2 groups. If the coxPH assumption is valid, then the base line functions from each group in the stratified PH model should be proportional

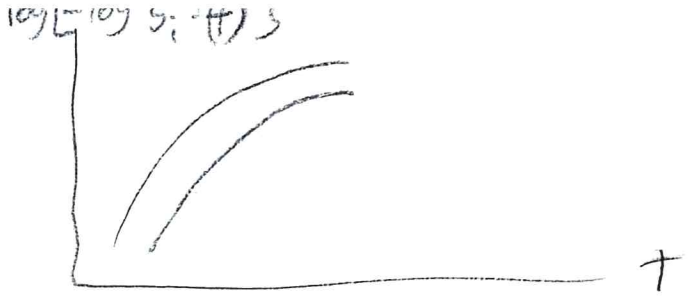
So $h_1(t) = d\, h_2(t)$ and $H_1(t) = d\, H_2(t)$

So $S_1(t) = \exp[-d\, H_2(t)] = [\exp -H_2(t)]^d = [S_2(t)]^d$,

So $\log[-\log S_1(t)] = -d \log S_2(t)$ and

$\log[-\log S_1(t)] = \log d + \log(-\log S_2(t))$.

Hence the $\log(-\log \hat{S}_1(t))$ and $\log(-\log \hat{S}_2(t))$ curves should be roughly parallel if the cox PH assumption is valid.



8] The ~~stratified~~ PH model works if the categorical (strata) covariate interacts with time. The particular form of the interaction (X * time, X * log(time) etc) does not need to be known,

9] If the form of interaction is known, other methods ~~of ch~~ are more efficient (estimators have smaller standard errors, tests are more precise) than the stratified PH model.

10] Allison p161 Stratification can be useful if there are clusters of cases such that observations within the clusters are not independent.

ex] children within schools: stratify on school
patients within hospitals (study sites): stratify on hospital

SAS variable selection (p.) but seemed to ignore the strata statement in that I got the same results without the strata statement (using recid data strata fin as in Allison p 161).

11] could stratify on a continuous variable by grouping the variable and using the groups as categori

If the categorical variable is fixed by the design of the study, it may be useful to use that variable as a strata variable.

ex) often sites (eg hospitals) are fixed by the design rather than a random sample of sites

13) You can still use the cox.zph function (that plots martingale residuals vs x_j with loess added) to assess the PH assumption for stratified PH models. The loess curves should be approximately horizontal.

14) There are lots of variants. SAS allows you to stratify on the variable and include it as a regular variable.

15) In R, you can stratify on a variable for the Weibull model. Again, several variants are possible.

~~Ch8~~ §7.3

1) In the PH model, predictors x_j are not allowed to depend on time.

2) P251 In the generalized Cox regression OCR model, the predictors $x_j(t)$ do depend on time. These predictors are called time dependent variables.

3) p251 Often patients are monitored for the duration of the study and some variables are recorded on a regular basis.

ex) site of tumoor
PSA levels for prostate cancer
white blood cell count
weight

external often only needs one initial measurement

§8.1 4) There are 2 types of time dependent variables. An internal time dependent variable is subject specific and requires the subject be under periodic observation. An external time dependent variable does not require the subject to be under direct observation.

ex) presence of side effects internal
$X_j * \log(\text{time})$ interaction external

age measured yearly external
environmental variables such as pollen count external
serum cholesterol level measured monthly internal

white blood cell count measured monthly internal

variance
age is known at start of study, you can compute only time afterwards

5) If $X_j(t)$ is the value of X_j measured at time t, the time t depends on study time not calendar time.

ex) If subject 1 began on May 1 and subject 2 on July 1 and both are measured weekly, then

the time in days will be 1, 14, 21 etc.

6) p253 * Let $\underline{x}_i(t) = (x_{i1}(t), \ldots, x_{ip}(t))'$

If x_j is not a time dependent variable, interpret

$x_j(t) \equiv x_j(0) = x_j$ so $x_{ij}(t) \equiv x_{ij}(0)$,

Then the generalized Cox regression model has

$\lambda(t) = \lambda_i(t) = \lambda(t, \underline{x}_i(t)) = \exp(\underline{\beta}' \underline{x}_i(t))\, h_0(t)$. Note that

$\underline{\beta}$ does not depend on t. The GCR model is not a PH model, but $h_0(t)$ is still the baseline function. The baseline function is for people who have $\underline{x}_i(t) \equiv \underline{x}_i(0) = \underline{0}$, $\forall t > 0$, that is, their variables are 0 at the time of origin and remain 0 through time.

ex) Suppose hazard of arrest depends on employment status. Could define

$x(t) = \begin{cases} 1 & \text{employed at time } t \\ 0 & \text{else} \end{cases}$

$x(t) = \begin{cases} 1 & \text{employed in previous month at time } t \\ 0 & \text{else} \end{cases}$

$x(t) =$ # of weeks of employment in the preceding 3 months at time t

$x(t) =$ # of bouts of unemployment in the preceding 12 months at time t

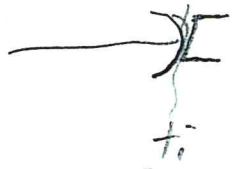
7) p253 Note that $\frac{h_i(t)}{h_0(t)} = \exp(\underline{\beta}' \underline{x}_i(t))$ depends on time. Also $h_i(t) \neq C\, h_0(t)$. So the GCR model is not a PH model.

(C: constant w.r.t. t)

8) p253 Data management and computing the GCR model is much more difficult than that for the Cox PH model.

9) For computation of the GCR model $\underline{x}_j(t_i)$ needs to be known for all individuals who are in the risk set at time t_i for $i = 1, \ldots, m$ (or there are missing values)

Recall that the t_i are the ordered ~~censoring~~ death times and the risk set contains all individuals who have not died or been censored until time t_i or later.



10) One type of time dependent covariate that is easy to work with is an interaction like X * time or X * log(time)

11) Application: Suppose a Cox PH model with variables $X_1, \ldots, X_p$ is fit. To test the Cox PH assumption, add variables $X_1 * \log(\text{time}), \ldots, X_p * \log(\text{time})$, fit a GCR model. Want the p values for the

interactions to be larger than .05 (7/21

(but multiple testing so with 20 interactions, all with H_0 $\beta_{1time} = 0$, about 1 in 20 will have pval < .05)

SAS program Test full vs reduced without interactions.

ex) proc phreg data = example; ← semicolon

model time*status(0) = x1 x2 x1lt x2lt

/ties = efron;

x1lt = x1 * log(time);

x2lt = x2 * log(time);

run;

GCR output

reduced model { x1, x2

x1lt

x2lt

Test reduced model.

	pval	
	.0001	
	.0798	
	.5123	} suggests co
	.0738	} PH assumptio is reasonable

See HW 11 #1 for an example using the recid data

ex) Stanford Heart Transplant data

Patients were enrolled in the study from 1967–1974 and followed until death or April 1, 1974.

103 patients: 69 received transplants (24 censored, 45 died); 34 did not (4 censored, 30 died)

Y = time from date of acceptance until death

age accpt = age at admission

surg = 1

time of transplant

surg = 1 if open heart surgery prior to acceptance 0 else

MI = # donor alleles with no match in recipient (1–4)

$M2 = \begin{cases} 1 & \text{if donor and recipient mismatch on HLA-A2 antigen} \\ 0 & \text{else} \end{cases}$

M3 = mismatch score

time of transplant, M1, M2, M3 were coded as missing if patient did not have transplant

$trans = \begin{cases} 1 & \text{had transplant} \\ 0 & \text{else} \end{cases}$

Cox PH model

```
proc phreg data = stan3 options yearcutoff=1900;
    model surv1*dead(0) = trans surg ageacept
```

test

-2 log L 45.463 with 3 df p = .0001

variable	par est. mate	SE	wald chi-sq.	Pr > chi-sq
trans	-1.7081	.2786	37.59048	.0001
surg	-0.4214	.3710	1.2902	.2560
ageacept	0.0586	.0151	15.1631	.0001

Problem with model: The main reason patients did not get a transplant was that they died before a suitable donor could be found. So death rates $\frac{30}{34}$ vs $\frac{45}{69}$ are much higher for those who did not get transplants. The predictor trans is a consequence of Y; early death prevents a patient from getting a transplant.

Let wait = time from admission until transplant

Let dot = time of transplant.

Let surv2 = time from transplant until death.

Could restrict analysis to those who received a transplant

```
proc phreg data = stan3;
where trans = 1;
model surv2*dead(0) = surg m1 m2 m3 agetrans wait dot
```

-2 log L 16.586 with 7 df pval = .0203

	parameter estimate	stand err	wald chisq	Pr > chisq
surg	-.7703	.4972	2.4004	.1213
m1	-.2486	.1944	1.6355	.2009
m2	.0296	.4427	.0045	.9467
m3	.6441	.3428	3.5309	.0602
agetrans	.0493	.0228	4.6619	.0308
wait	-.0020	.0051	.1469	.7015
dot	-.0002	.0003	.3044	.5811

$\beta = 0$ would not be rejected at $\alpha = .02$

older patients have higher risk of dying

ex) Ex A had fixed covariate trans = $\begin{cases} 1 & \text{patient had transp} \\ 0 & \text{else} \end{cases}$

and results were misleading since patients who die quickly are less likely to get a transplant (making the transplant procedure seem much more effective than it actually is).

In Ex B can't compare those who got a transplant to those who did not which is the main goal of the study.

Let plant $= X_i(t) = \begin{cases} 1 & \text{patient has had transplant by time } t \\ 0 & \text{else} \end{cases}$

options yearcutoff=1900;
proc phreg data=stan;
model surv1*dead(0)=plant surg ageaccpt / ties=exact;
if wait>surv1 or wait=. then plant=0; else plant=1;
run;

surv1 = time until death surg = 1 : if patient has had previous heart surgery

ageaccpt = age until transplant (coded as missing . for those who did not receive transplants) wait = time (from acceptance)

The if condition is true if wait is missing and surv1 will be ti of some patient who died at the beginning of the risk set.

	estimate	SE	chisq	Pr>chisq ← from Allison
plant	-0.0462	0.3028	0.0232	.8788
surg	-0.7715	0.3596	4.6022	.0319
ageaccpt	0.0311	0.0139	4.9952	.0254

I did not get these values from SAS

Transplantation seems to have no effect on hazard of death

ex D) To make age a time varying covariate use

options yearcutoff=1900; ← add this to the data set and SAS? Ingersoll with Allison. I have a Y2K problem.
proc phreg data=stan;
model surv1*dead(0) = plant surg age / ties=exact;
if wait>surv1 or wait=. then plant=0; else plant=1;
age = ageaccpt + surv1;
run;

wrong output Y2K bug

	estimate	SE	chisq	Pr>chisq
plant	0.00387	0.2976	0.0002	0.9896
surg	-0.40526	0.3784	1.5121	0.2188
age	-0.00856	0.00792	0.5620	0.4534

← from SAS

$h_{x+1}(t) = \exp(\beta' x(t))\, h_0(t)$

so $\log\left(h_{x+1}(t)\right) = \log h_0(t) + \beta_1 x_1(t) + \beta_2 x_2 + \beta_3 x_3(t)$

(prent, surg, age)

$x_3(t) = x_3(0) + t = x_3 + t$

$= \log h_0^*(t) + \beta_1 x_1(t) + \beta_2 x_2 + \beta_3 x_3$

$\log h_0(t) + t = \log h_0(t) + \log e^t = \log e^t h_0(t)$ so $h_0^*(t) = e^t h_0(t)$

So changing x_3 to $x_3(t)$ should not affect output of exC and D. I took exC output from Allison and exD from SAS. If both are taken from SAS, they agree.

~~p347 gives another SAS ex~~

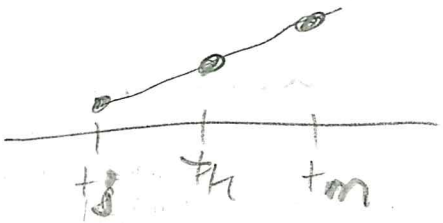
12) p255 $X_{ij}(t)$ needs to be known at death

i = ith case, j = jth predictor

time t_k if case i is in the risk set at time t_k. Often $X_{ij}(t_k)$ is approximated.

a) Use last recorded value of $X_{ij}(t)$ before time t_k.

b) Use linear interpolation



§8.3 p258

13) Know for final Inference for the GCR model is almost exactly the same as inference for the PH model. See ex2 review 21) – 24).

But in 22) and 23), replace "PH" by "GLR" in the conclusion.

14] P257 For the GCR model,

$$S_{\underline{X}(t)}(t+1) \neq \{\bar{S}_0(t+1)\}^{e^{\underline{\beta}'\underline{x}(t+1)}}.$$

$$S_{\underline{X}(t+1)}(t+1) = \exp\left[-\int_0^t \exp(\underline{\beta}'\underline{x}(u))\, h_0(u)\, du\right]$$

So $\underline{x}(u)$ needs to be known on $(0,t)$ and may depend on unknown future values.

It is possible to estimate $P(Y_i > t+\Delta \mid Y_i \geq t)$

15] P259 $\underline{\beta} = \underline{0}$ means changing $\underline{x} = \underline{x}(t)$, within the range of observed $\underline{x}$, does not change survival.

$\beta_i = 0$ means changing $x_i = x_i(t)$, within the range of observed x_i, does not change survival.

$\underline{\beta} = (\underline{\beta}_R', \underline{\beta}_0')'$ and $\underline{\beta}_0 = \underline{0}$ (reduced model is good) means changing values of $\underline{x}_0$, within the range of observed values of $\underline{x}_0$, does not change survival.

So "no survival relationship" between Y and $\underline{x}$ or x_i or $\underline{x}_0$ means within the observed range of $\underline{x}$ or x_i or $\underline{x}_0$

ex] $X_i = trt = \begin{cases} 0 & trt\ 0 \\ 1 & trt\ 1 \end{cases}$ and trt 1 and trt 0 both good but equally effective (no treatment effect) means $\beta_i = 0$ since $e^{\beta_i x_1}\, e^{\sum_{i=2}^{p} \beta_i x_i}\, h_0(t)$ needs to be the same for $x_1 = 0$ and $x_1 = 1$.

But if no treatment is given, eg $x_1 = 2$, survival could decrease

dramatically.

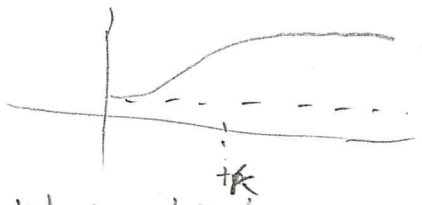
Similarly $h_{\underline{x}}(t) = e^{\underline{\beta}_R' \underline{x}_R} h_0(t) = e^{\underline{\beta}_0' \underline{x}_0} e^{\underline{\beta}_R' \underline{x}_R} h_0(t)$

$\Rightarrow \underline{\beta}_0 = \underline{0}$. But if X_1 is a variable in $\underline{x}_0$ and no trt is given ($X_1 \to X_1 = 2$, oct of the observed range of X_1), then survival could decrease dramatically

16] p259 Time dependent variables could change due to the treatment X_1, with $\hat{\beta}_1$ such that Ho $\beta_1 = 0$ not rejected even though X_1 is important.

ex) p259 Suppose trt $= X_1 = \begin{cases} 0 & \text{placebo} = \text{sham treatment} \\ 1 & \text{medicine} \end{cases}$

Survival of patients with leukemia may depend on white blood cell count $= X_2$.



So $X_1 = 1$ means $X_2(t)$ is high after time t_K and survival is high while $X_1 = 0$ means $X_2(t)$ is low for all t and survival is low. But X_1 may not be needed in the GCR model (fail to reject Ho $\beta_1 = 0$) given $X_2(t)$ is in the model. Then the time dependent covariate $X_2(t)$ masked the treatment effect. Basically $X_2(t)$ has accounted for the treatment difference and explains why the treatment is effective. Note that Ho $\underline{\beta} = \underline{0}$ would be rejected since there is a change in survival as $\underline{x}$ values change.

17} p 259 The martingale residual for the ith subject is $\hat{r}_{Mi} = \delta_i - \exp\left(\hat{\beta}' X_i(y_i^*)\right) \hat{H}_0(y_i^*) \in (-\infty, 1]$

$e^{\hat{\beta}' X_i} \hat{H}_0 \geq 0$

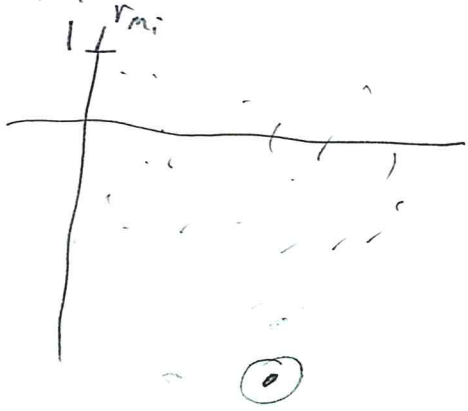
where $\delta_i = \begin{cases} 0 & y_i^* = z_i \text{ is censored} \\ 1 & y_i^* = y_i \text{ is uncensored} \end{cases}$

The deviance residual $\hat{r}_{Di} = \text{sgn}(\hat{r}_{Mi}) \sqrt{-2[\hat{r}_{Mi} + \delta_i \log(\delta_i - \hat{r}_{Mi})]}$

where $\text{sgn}(r_{Mi}) = \begin{cases} 1 & r_{Mi} > 0 \\ -1 & r_{Mi} < 0 \end{cases}$

18} For GCR) Plotting martingale residuals vs X_i, adding loess, and checking that loess is roughly linear is not nearly as useful for checking whether $w_i = t(X_i)$ should be used instead of X_i

19} plot i vs r_{Mi} to detect outliers



§8.4

20} p 260 Collett suggests using X_i + interactions to test the PH assumption. We will use $X_i \log(t)$ interactions.

ex} Heart attack Y = survival time
$X(t)$ = time from heart attack to arrival at intensive care unit is an important predictor

ex} p 261 $X_1(t) = \begin{cases} 1 & \text{transplant by time } t \\ 0 & \text{else} \end{cases}$

$h_{X(t)}(t) = \exp(\beta_1 X_1(t) + \beta_2 x_2 + \cdots + \beta_p x_p) h_0(t)$

In PH model $\underline{\beta} = \underline{0}$ does not necessarily mean $\underline{X}$ is unimportant. Suppose $p=1$

$X = \begin{cases} 1 & \text{trt 1} \\ 0 & \text{trt 0} \end{cases}$. If trt 1 & 0 are both very but equally effective

$h_1(t) = \lambda_0(t) = e^{\beta X} h_0(t)$ so $\beta = 0$.

So $\underline{\beta} = \underline{0}$ could imply the survival relationship between Y and $\underline{X}$ (is the same (as that between Y and $\underline{X} = 0$ if 0 is a reasonable value of $\underline{X}$) for all observed values of $\underline{X}$.

If $\log Y = \alpha + \beta' x + e$ can get same interpretation.

Note that $trt = \begin{cases} 1 & \text{trt 1} \\ 2 & \text{trt 2} \end{cases}$ still has $\beta_{trt} = 0$ if trt 1 & 2 are equally effective. Then $\frac{h_1(t)}{h_2(t)} \equiv 1$.

$\underline{\beta} = 0$ means changing values of $\underline{X}$ does not effect survival (within the range of observed values of $\underline{X}$ or $\underline{\beta}'\underline{X}$)

$\beta_i = 0$ means changing values of x_i does not effect survival (within the range of observed values)

$\underline{\beta} = (\underline{\beta}_R', \beta_0')'$ and $\beta_0 = 0$ means changing values of x_0 within the range of observed values of x_0 does not effect survival

or $= 0$

So reduced model is good in that you get the "same survival model" regardless of the x_0 values

So "no survival relationship" between Y and $\underline{X}$ or x_0 or x_i means within the observed range of $\underline{X}$ or x_0 or x_i

ignore

YARD DEBRIS SERVICE

MADE EASY

CALL

521-DBRI

521-3274

½ pickup load $11.00
pickup load --$20.00
7' X 16 X 2 trailer load $36.00

Loads may consist of leaves, branches, metal, old broken fencing, --- all at the same low price.

Other custom hauling available.
Driveway rock delivered @ $32.00 a ton – 2 ton minimum

Trees available for planting, put your order in for fall delivery.

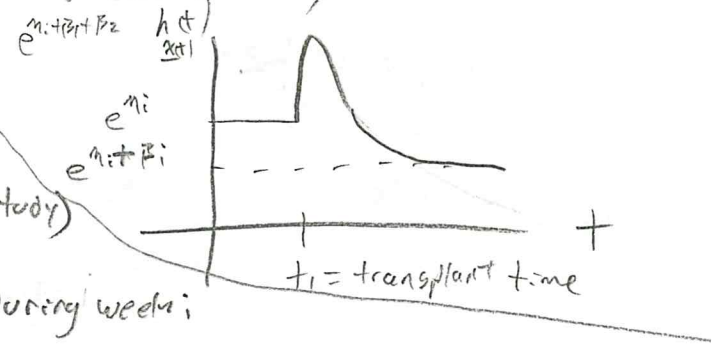
So $h_{\underline{x}(t)}(t) = \begin{cases} \exp(\beta_2 x_2 + \cdots + \beta_p x_p) h_0(t), & \text{no transplant by time } t \\ e^{\beta_1} \exp(\beta_2 x_2 + \cdots + \beta_p x_p) h_0(t), & \text{transplant by time } t \end{cases}$

For 2 patients with the same $x_2, \ldots, x_p$

$$\frac{h_{\underline{x}(t)}(t \mid x_1(t)=1)}{h_{\underline{x}(t)}(t \mid x_1(t)=0)} = e^{\beta_1} \qquad \text{so } \beta_1 > 0 \Rightarrow$$

transplant increases hazard while $\beta_1 < 0 \Rightarrow$ transplant decreases hazard. Can do a wald test for β_1 or test the reduced model that does not contain $x_1(t)$.

2) often hazard increases immediately after a transplant then decreases.



ex) The recid data (year long study)

has empi = $\begin{cases} 1 & \text{employed full time during week } i \\ 0 & \text{else} \end{cases}$

for $i = 1, \ldots, 52$. Let $X_8(t) = \begin{cases} 1 & \text{employed during week } i \\ 0 \end{cases}$

Time until arrested was measured in weeks so $X_8(t)$ is time dependent.

$\hat{\beta}_8 = -1.328321 \qquad X_8^2 = 28.07006 \qquad p\text{-val} = .0001$

Problem arrests may affect employment status rather than vice versa: if a subject is arrested at the beginning of the week, the prob of working full time that week drops a lot

Could use $x_8(t) = \begin{cases} 1 & \text{employed full time the previous week} \\ 0 & \text{else} \end{cases}$

Then $\hat{\beta}_8 \simeq -0.79$ is roughly halved but $x_8^2 = 13.1$ has pval $< .001$.

§8.5 ex 8.1 p261 predictors group $\begin{cases} 1 & \text{AL (acute lymphoblastic) leukemia} \\ 2 & \text{low risk AM leukemia} \\ 3 & \text{high risk AM leukemia} \end{cases}$

page = age of patient
dage = age of bone marrow donor

$p = \begin{cases} 1 & \text{blood platelet count returned to normal} \\ 0 & \text{else} \end{cases}$

ptime = time for platelet count to return to normal if $p=1$

$\text{plate}(t) = \begin{cases} 0 & \text{if } t \le \text{time at which count returned to normal} \\ 1 & \text{if } t > \text{" "} \end{cases}$

null model $-2\log L = 67.13$

model with plate(t) $-2\log L = 62.21$ pval = .026

x_1

$\hat{\beta}_1 = -2.696$ so hazard is higher for a patient whose platelet count is not normal

model with plate(t) group ind 1 group ind 2 had $-2\log L = 55.72$
x_1 x_2 x_3

pvalue for group ($X^2(\text{LR}) = 6.49$ on 2 df) $= .039$

ex 8.2 p265 ovarian cancer data predictors

age, treat, age * time

time dependent variable

$\hat{\beta}_3 = .0002$ pval $= .465$

so PH assumption is not violated

(Usually $X_i * \log(\text{time})$ is used to check the PH assumption)

ex 8.3 p 266 Artificial data to illustrate a time dependent variable that arises from multiple measurements on each patient. This type of data set is usually very large.

Y = survival time in days

treat = $\begin{cases} 0 & \text{placebo} \\ 1 & \text{liverol} \end{cases}$

age

Lbrt = log bilirubin level, measured 0, 3, 6, 12 (roughly) months after commencement of treatment, then "yearly". Table 8.3 gives Lbr at time 0 and table 8.4 for other times in days.

So Lbr is measured at t = 0, 47, 184, 251 for patient 1 and at t = 0, 74, 202, 346, 917, 1411 for patient 7. Lbr is treated as a step function so Lbrt(ti) is the most recent Lbr measurement. Lbr = Lbrt(0) is a fixed covariate.

PH	-2log L
Null	25.121
Age	22.135
Lbr	21.662
Age, Lbr	18.475
Age Lbr treat	13.293

pval = .1023 for treat

β_3 = -3.052 so liverol is effective

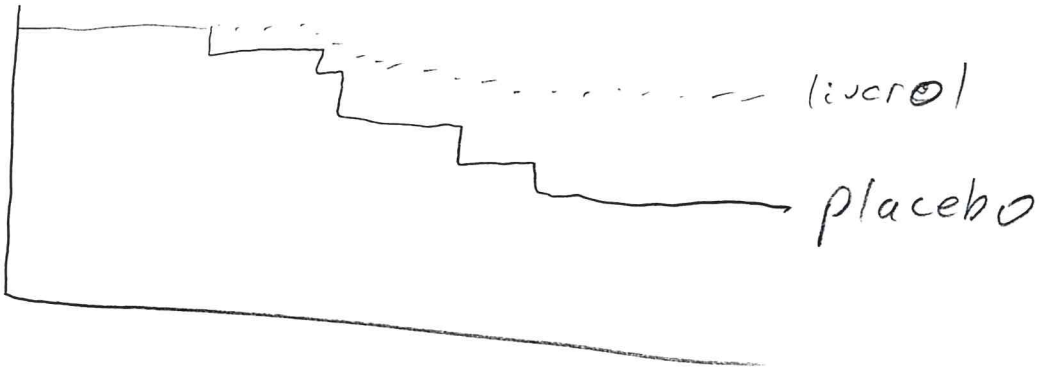
GBR	-2log L
Null	25.121
Age	22.135
Lbrt	12.050
Age Lbrt	11.145
Lbrt, treat (β_1 β_2)	10.676

pval for treat = .241

Lbrt could be masking the treatment effect

higher Lbr(t) values seem better

Fig 8.5 gives baseline survival functions, and patients on placebo do worse



PH model

Shortcut in R for Ho reduced model is good

full

> fit ← coxph(Surv (time, status) ~ resid.ds * rx + ecog.ps, data = ovarian)

> fit.R ← coxph(Surv time, status) ~ resid.ds + rx + ecog.ps , data = ovarian)

reduced

	coef	exp(coef)	SE(coef)	z	p
full			2.326	1.79	0.073
resid.ds	4.236		2.556	1.05	0.296
rx	2.674		0.592	0.80	0.424
ecog.ps	0.473		1.421	-1.38	0.167
resid.ds:rx	-1.936				

LRT = 6.08 on 4 df P = 0.0888

reduced				
resid.ds	1.347	0.680	1.98	0.048
rx	-0.749	0.595	-1.26	0.208
ecog.ps	0.455	0.590	0.77	0.443

LRT = 6.03 on 3 df P = 0.111

> anova(fit2, fit, test = "Chisq")

	loglik	Chisq	Df	P(>|Chi|)
1	-31.970			
2	-30.944	2.0469,	1	0.1525

χ^2 DF=1 — 0.05 — 0.01 — Dual — so reduced model is good