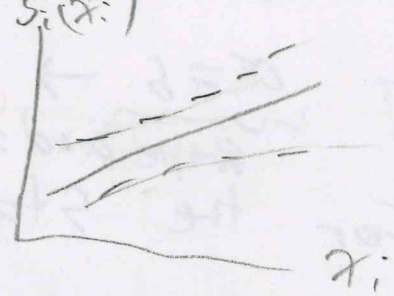
4) Plot x_i vs $\hat{S}_i(x_i)$. If the plot is linear, use x_i in the PH model.



If the plot is quadratic, add x_i^2 to the PH model.

If the plot is cubic, add x_i^2 and x_i^3 to the model.

Let $w_1, \ldots, w_k$ be the final set of predictors in the PH model eg $\{x_1, x_1^2, x_2, x_3, x_4, x_4^2, x_4^3\}$

$w_1 \; w_2 \; w_3 \; w_4 \; w_5 \; w_6 \; w_7$

Fit the GAM to this model and check that all w_i vs $\hat{S}_i(w_i)$ plots are linear.

(from Wood p 116–119, 283–287)

Another AFT

13) Another AFT is the generalized gamma model. Then there are intercept α, scale σ, and shape δ parameters.

intercept
x_1
$\vdots$
x_p
scale
shape

If $\delta = 0 \rightarrow$ lognormal

$\delta = 1 \rightarrow$ weibull $\rightarrow (\sigma = 1, \delta = 1 \rightarrow$ EXP$)$

$\underbrace{\sigma = \delta}_{\text{shape and scale parameters are equal.}} \rightarrow$ standard Gamma model (2 parameter)

For the standard Gamma model, the pdf of Y

is $$f(t) = \frac{\lambda (\lambda t)^{k-1} e^{-\lambda t}}{\Gamma(k)}, \quad k = \frac{1}{\delta^2} > 0$$

and $\lambda = \exp[-(\alpha + \beta^T x)] > 0$.

Note that $$f(t) = \frac{t^{k-1} e^{-\lambda t}}{\lambda^{-k} \Gamma(k)}$$

Some texts take $\lambda = \frac{1}{\beta}$.

$E[Y|x] = \frac{k}{\lambda}$, $V(Y|x) = \frac{k}{\lambda^2}$.

could make EE plots to test lognormal, weibull, EXP, or standard Gamma vs generalized Gamma.

Competing Risks

1) The event could be due to J causes of interest.

ex) death : heart disease, cancer, stroke, accident, other

473 64

2} Let $h_i(t)$ be the hazard function for the ith person with $\underline{x_i}$. Let $h_{ij}(t)$ be the hazard function for death type j for the ith person.

Then $h_i(t) = \sum_{j=1}^{J} h_{ij}(t)$.

3} PH: $\log h_{ij}(t) = \underbrace{\alpha_j(t)}_{\log h_{0j}(t)} + \underline{\beta_j}' \underline{x_j}$, $j = 1, \ldots, J$

4} Can fit all different models

eg PH for $j=1$, GCR, $j=2$, WAFT $j=3$

5} Can focus on type j by considering deaths by causes $i = 1, \ldots, j-1, j+1, \ldots, J$ to be censored for type j.

ex) 472 leaders from countries outside of North America, Europe and Australia

Y = # years in power and the

competing risks are the mode of exit from power

death from natural causes

constitutional means

nonconstitutional means such as assassination, overthrow

426 cases: 179 still leaders in 1987 0

(rise to power 1960-1987):

27 died of natural causes 2

115 constitutional 1

115 unconstitutional 3

lost = competing risks variable

variables manner leader reached power

0 constitutional

1 unconstitutional

start = year of entry into power

military 1 if military 0 if civilian

age (at start)

conflict = level of ethnic conflict 0 low 1 med or high

log inc log (GNP)

growth annual ave rate of percapita GNP growth between 1965 and 1983

pop

land (area)

literacy (rate)

region 0 - middle east 1 Africa 2 Asia 3 Latin America

model years * lost(0,1,2) = ... focuses on type 3 SAS

model years * lost(0,1,3) = ... 2

model years * $\underbrace{\text{lost (0,2,3)}}_{\text{treated as censored}}$ = ... 1

The output for the 3 types of competing risks is analyzed like output for PH data

$\hat{\underset{\sim}{\beta}}_1$ type 1 SE wald chisq pr > chisq

x1
⋮
xp

$\hat{\underset{\sim}{\beta}}_2$ type 2

x1
⋮
xp

$\hat{\underset{\sim}{\beta}}_3$ type J=3

x1
⋮
xp

6) If $\underset{\sim}{\beta}_1 = \cdots = \underset{\sim}{\beta}_J$ and PH was used,

just use the PH model based on all of the data (competing risks model is not useful and $h_0(t) = \sum_j h_{0j}(t)$)

7) Problem: a) times for different event types need to be independent, or at least each type is noninformative for the others.

b) Often death is of much more interest than the competing types of death. Then just use a usual model such as PH.

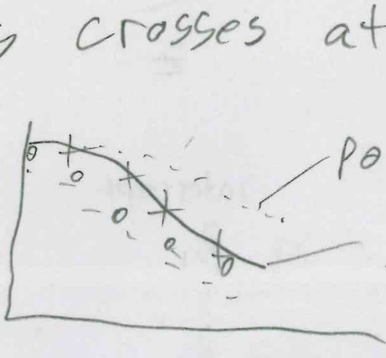
Weibull PH regression survival function is

$$\hat{S}_{w\underline{x}}(t) = \exp\left(-e^{ESP}\, \hat{\lambda}_0\, t^{\hat{\gamma}}\right) \text{ where}$$

$$ESP = \frac{-\hat{\beta}_A{}'\underline{x}}{\hat{\sigma}}, \quad \hat{\lambda}_0 = e^{-\hat{\gamma}\hat{\alpha}}, \quad \hat{\gamma} = \frac{1}{\hat{\sigma}}.$$

$$\log Y_i = \alpha + \beta_A' x + \sigma \varepsilon_i \quad (\text{weibull AFT}).$$

Since the weibull regression model is a PH model, make the EE plot. If the EE plot is good, modify the slice survival plot by adding the weibull $\hat{S}_{w\underline{x}}(t)$ as crosses at $Y_{(i)}^*$. Typically the crosses and solid line are very close.



pointwise CI bands

$\hat{S}_{c\underline{x}}(t)$ = solid line

o = KM

$+ = \hat{S}_{w\underline{x}}(Y_{(i)}^*)$

ordered observed times both censored and uncensored

See "plots for Survival Regression" on my preprints page

Final is cumulative

1) A] $F(t)$ $f(t)$ $S(t)$ $h(t)$ $H(t)$: usually 2 are easy 2 are harder. Do the harder ones in 2 ways as a check.

12) B) KM estimator $n_i = \sum_{j=1}^{n} I(t_{(j)} \geq t_i)$

d_i = # deaths at t_i, t_i is a death time

$t_{(j)}$ is a death or censoring time and contributes

to n_i but not to d_i. censored $t_{(j)}$ $\hat{S}_{KM}(t_i) = \hat{S}_{KM}(t_{i-1})\left(1 - \frac{d_i}{n_i}\right)$

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

If $t_M = t_{(n)}$ then $\hat{S}_K(t) = 0$ for $t > t_M$

If $t_{(n)} > t_M$, then $\hat{S}_K(t) = \hat{S}_K(t_M)$ for $t_M \leq t \leq t_{(n)}$ but

$\hat{S}_K(t)$ is undefined for $t > t_{(n)}$.

C)

PH
SPH
GCR

SAS gives $-2\log L$(full) X^2 = chisq

R Z

$X^2(N|F)$ from (partial) likelihood ratio test

($X^2(N|R)$ for reduced model)

WPH $\Leftarrow$ Weibull AFT iff $\log Y$ has

other AFT models

SAS X^2 = chisq

intercept
x_1 $\hat{\beta}_1$
⋮
x_p $\hat{\beta}_p$

for WPH $\{$ scale, weibull shape

Need $\log L$(Null)

gives $\log L$(full)

then $X^2(N|F) = [-2 \log L(\text{null})] - [-2 \log L(\text{full})]$

R Z

intercept
x_1 $\hat{\beta}_1$
⋮
x_p $\hat{\beta}_p$

for WPH $\{$ log scale

chisq $= X^2(N|F)$

($X^2(N|R)$ for reduced model)

Fred 2

D) plots

i) Slice survival plot for goodness of fit

$\hat{S}_{\underline{x}_j}(t)$ vs $\hat{S}_{KM_j}(t)$ for jth group found by dividing ESP into J groups. Add $\hat{S}_{W\underline{x}_j}(t)$ if WPH is used.

ii) ET plot of ESP vs T for PH

iii) EE plot of parametric ESP vs Cox ESP especially for WPH: looks like identity line

iv) LRC plot of ESP vs log T should be linear for AFT models: identity line in upper left corner, left skewed data

v) H plots of x_i vs Schoenfeld residuals loess should be horizontal if PH assumptions are reasonable GLOBAL pval > 0.05 ⇒

vi) L plots of x_i vs martingale residuals

if loess curve $g(x_i)$ is linear, use x_i

if loess $g(x_i)$ is nonlinear, replace x_i by $g(x_i)$ Here $g(x_i)$ should be a simple function like $h(x_i) = x_i^2$.

E other problems from Exams 1-3, GGR Q10 like the lifetable estimator are possible

E) Shorth

F) Bootstrap

G) $\underline{\hat{\beta}_{VS}} = \underline{\hat{\beta}}_{Imin}$

Quiz 10