EXAM C QUESTIONS OF THE WEEK

S. Broverman, 2006

Week of February 20/06

A Cox proportional hazards model is applied to model the future mortality of non-smokers and smokers who are currently 50 years old. The model assumes that non-smoker mortality follows DeMoivre's law with upper age limit 100, so that $h_0(t) = \frac{1}{50-t}$. There is a single covariate, Z, and $Z = 0$ indicates a non-smoker, and $Z = 1$ indicates a smoker. Age of death is available for 10 individuals now at age 50, 5 of whom are smokers and 5 of whom are non-smokers. The ages of death of the smokers are 55 , 58 , 62 , 67 , 77 , and the ages of death of the non-smokers are 58 , 63 , 72 , 81 , 91 . Find the maximum likelihood estimate of the probability that a 50 year smoker will live to at least age 75.

Solution can be found below.

Week of February 20/06 - Solution

With (parametric) baseline hazard function $h_0(t) = \frac{1}{50-t}$, the baseline pdf for time until death for a 50-year old non-smoker is $f_0(t) = \frac{1}{50}$ for $0 < t < 50$ (DeMoivre's Law corresponds to a uniform distribution for the remaining time until the upper age of the survival distribution - there are 50 years remaining until age 100 for the 50-year old). The smoker hazard rate will be $h_s(t) = c \cdot h_0(t) = \frac{c}{50-t}$, and the pdf of a 50-year old smoker's time until death is $f_s(t) = f_0(t) \cdot c \cdot [S_0(t)]^{c-1} = \frac{c}{50} \cdot (\frac{50-t}{50})^{c-1}$.

The log-density for a non-smoker is $ln\, f_0(t) = -ln\, 50$ for any death time t, and the log-density of death at time t of a 50-year old smoker is

$ln\, f_s(t) = ln\, c - ln\, 50 + (c-1)ln(50-t) - (c-1)ln\, 50$.

For the given ages of death, the death times (measured from age 50) for the smokers are 5 , 8 , 12 , 17 , 27 , and for the non-smokers are 8 , 13 , 22 , 31 , 41 .

The loglikelihood ℓ for the data set is the sum of the log densities, which is

$\ell = 5 \cdot (-ln\, 50) + 5\, ln\, c - 5 \cdot (ln\, 50) + (c-1)\sum_S ln(50-t_i) - 5(c-1)\, ln\, 50$,

where the summation is over the smoker's death times.

To find the maximum likelihood estimate of c, we set $\frac{d}{dc}\ell = 0$, so that

$\frac{5}{c} + \sum_S ln(50-t_i) - 5\, ln\, 50 = 0$.

Using the 5 smoker's death times, we have

$\frac{5}{c} + ln(50-5) + ln(50-8) + ln(50-12) + ln(50-17) + ln(50-27) - 5ln\, 50 = 0$.

Solving for c results in $c = 2.863$.

Then, the survival probability for t years from age for a smoker is $S_s(t) = [S_0(t)]^{2.863}$, so that

$S_s(25) = [S_0(25)]^{2.863} = (\frac{50-25}{50})^{2.863} = .14$.