MTH6157 Survival Models

January 2021 main exam Solutions

Question 1

(a) at week j let n_j be the number of students still active in the module at the beginning of the week, d_j the number who fail the quiz that week and c_j the number right censored by not joining the online lecture for week j+1

NOTE Kaplan Meier assumes censoring occurs after decrements therefore those not logging on in week j need to be counted after those that fail the test in week j-1 so are c_{j-1} not c_j [this is a difficult point testing knowledge of Kaplan Meier]

then the hazard at week j is λ_j = dj / nj and the Kaplan Meier estimate of the survival function at week j is S(j) = $\prod_{i \ll j}(1 - \lambda i\ )$

We seek the largest j for which S(j) > 0.65

j	n	d	c	λ	1- λ	S(j)
0	20					1
1	20	1	1	0.05	0.95	0.95
2	18	1	0	0.0556	0.9444	0.8972
3	17	0	0	0	1	0.8972
4	17	1	0	0.0588	0.9412	0.8444
5	16	2	3	0.125	0.875	0.7389
6	11	1	0	0.0910	0.9091	0.6717
7	10	2	1	0.2	0.8	0.5374
8	7	2	0	0.2857	0.7143	0.3838
9	5	0	0	0	1	0.3838
10	5	1	0*	0.2	0.8	0.3071

(*) could argue that the remaining students after the week 10 quiz are right censored by the end of the module so acceptable to have c_{10} as 0 or 4

By this Kaplan Meier estimate of the survival function we would set the pass criteria as successful completion of the week 6 quiz.

(b) assumptions:
- where no failed quiz is observed the hazard is zero
- that failure to log on is the only form of censoring
- that censoring occurs after quiz failure
- that it is reasonable to assign week j not login students to c in week j-1
- that the 20 sample students are representative of future year groups
- that future quizzes are of comparable standard

(c) concerns:

- the hazard function has an unusual shape rising then falling repeatedly
- relatively small sample size and large amount of censoring relative to failures
- what is leading to relatively large c_5? A lot (3 of remaining 14) not logging in to week 6 lecture
- what effect will knowledge of the pass criteria have on student behaviour
- the assumptions in (b) seem quite unrealistic, e.g. year to year comparability

Question 2

(a) baseline applies when all $z_i = 0$ so patient age < 65 with no other pre-existing lung conditions who last tested positive 2 months ago.

(b) $z_1=1$, $z_2=5$, $z_3=1$ therefore $(\beta.z^T) = (1\text{x}0.05)+(5\text{x}0.1)+(1\text{x}{-0.3}) = 0.25$

(i) $\lambda(t) = \lambda_0(t)\exp(0.25)$

(ii) $S(t) = \exp -\int_0^t \lambda 0(\mathrm{s})\exp(0.25)ds$

(c) for the 47 year old $z_1=0$, $z_2=1$, $z_3=0$ therefore $(\beta.z^T) = (0\text{x}0.05)+(1\text{x}0.1)+(0\text{x}{-0.3}) = 0.1$

so $S_{47}(4) = 0.93 = \exp - \int_0^4 \lambda 0(\mathrm{s})\exp\,(0.25)ds$

therefore, $-\int_0^4 \lambda 0(\mathrm{s})ds = \ln 0.93 / \exp 0.1$

and from (b)(ii) $S_{85}(4) = \exp -\int_0^4 \lambda 0(\mathrm{s})\exp(0.25)ds = \exp\,[\ln 0.93 \exp 0.25 / \exp 0.1]$
$= 0.919$

Question 3

(a) under the Poisson model, if D is the number of deaths at age x then

$P[D=d] = \exp(-\mu V).(\mu V)^d / d!$

where μ is the (constant) force of mortality at age x we seek to estimate

if first death is after time v_1, second death after v_2,…, k^{th} death after v_k,… the likelihood function is the product of the d individual Poisson probabilities at times v_k

$L(\mu) = \prod_{all\ k} \exp(-\mu.vk)\,.\,(\mu.vk)^k/k!$

and the log likelihood is

$\log L(\mu) = \sum_{all\ k}[k \log \mu.vk - \ \mu.vk - logk!]$

we differentiate and set to zero to find the MLE for μ

$d/d\mu \log L(\mu) = d/\mu - V = 0$ when $\mu = d/V$ where V is the sum of the v_k's for $1 \ll k \ll d$

so the Poisson MLE for μ here is d/V

(b) assumptions:
- a constant force of mortality within each calendar year of age
- a Poisson type distribution function represents the underlying mortality
- central exposed to risk can be accurately estimated
- the group of lives under consideration can be thought of as sample from the same underlying distribution for future lifetime

(c) these are most problematic at ages when:
- force of mortality is most rapidly accelerating (from older middle age onwards)
- there is very little data (very oldest ages 90+)
- there are reasons to sub-divide the population into smaller more homogenous groups based on some covariates
- and potentially for treatment of the male 'accident hump' in late teens / early 20s

(d) the company needs models which allow for multiple decrements (ill health, death, maybe others) and the Poisson model cannot easily be extended to multiple decrements. The company probably needs to switch to a multi-state model, but may be able to use the Poisson derived forces of mortality as the transition intensities to death. Health related transition intensities would need to be estimated separately.

Question 4

(a) rate of penalties per person hour = {number penalties} / {exposed to risk in hours}

number penalties = 200/50 = 4

use the Census method to approximate exposed to risk.

London – Milton Keynes	139 people x 30 minutes = 4170 minutes
Milton K – Stoke on T	(139+33-29) x 58 = 8294
Stoke on T – Stockport	(143+17-3) x 28 = 4396
Stockport – Manchester	(157+7-2) x 11 = 1782

total exposed to risk = 18642 minutes

rate per person hour = 4 / (18642/60) = 0.012874

(b) assumptions:
- the calculation itself is precise, it is the application of the rate to any analysis that requires assumptions
- if all 4 penalty people travelling together, or
- if all 4 penalties issued in same part of the journey
- then smoothed rate across the journey has less meaning

Question 5

(a) The null hypothesis is that the standard table is a good representation of the mortality experience. We check overall fit with a chi-squared test.

the chi-squared statistic is $z_x^2 = \sum \frac{(O-E)^2}{E}$ and with 8 ages is compared to $\chi^2_{0.95;8}$

where O = observed deaths and E = expected deaths under the null hypothesis

Age x	Observed	E-to-R	qx	Expected	zx^2
53	4	1387	0.0024	3.3288	0.1353
54	4	1348	0.0028	3.7744	0.0135
55	4	1304	0.0032	4.1728	0.0072
56	5	1294	0.0037	4.7878	0.0094
57	7	1283	0.0043	5.5169	0.3987
58	6	1263	0.0050	6.3150	0.0157
59	9	1238	0.0058	7.1804	0.4611
60	9	1203	0.0068	8.1804	0.0821
					1.1230

$\chi^2_{0.95;8}$ = 15.51 > 1.123 therefore we do NOT reject the null hypothesis

based on this test the table is a good representation of mortality experience

(b) the chi-squared test will not detect:
- an overall small increase in mortality rates due to the pandemic
- a temporary change in mortality rates are possible in this situation
- time selection – e.g. long Covid effects – not taken into account
- sample issues, 1 additional death in most ages would have considerable effect on z

NOTE the generic lecture note list of what chi-sq test does not detect is not enough here, need to link to question wording and pandemic context for marks

(c) further tests:
- signs test would show overall increase in mortality
- grouping of signs / Steven's test would indicate a correlated set of changes in mortality experience

(d) why particular concern:
- annuitants at these ages with expectation of future life > 30 years can be large financial liability for the insurer
- to manage financial risk, insurer would rather underestimate mortality rates than overestimate them
- the table is historic
- the test is based on existing policyholders not the new early retiring cohort
- possibility of adverse selection (early retirees being healthier than average)
- underwriting practices were probably designed in different circumstances