MTH6157 2021-22 Solutions

January paper

Q1 Survival Model Concepts

(a) those who die between ages 65 and 68 will not receive any pension

(b) either 100,000 ${}_3p_{65}$ or accept 100,000 $S_{65}(3)$

(c) as ${}_tp_x = \exp[- \int_0^t \mu_{x+s}\, ds]$ we can re-write (b) as ${}_3p_{65} = \exp[- \int_0^3 \mu_{65+s}\, ds]$

(d) occurrence exposure rate for a population is number of deaths / total observed time alive. Occurrence exposure rate is used to estimate the central rate of mortality (m_x) where,

$m_x = q_x / \int_0^1 {}_tp_x\, dt$

when the force of mortality is a constant μ then $m_x = \mu$ so that the occurrence exposure rate estimate for m_{65} m_{66} m_{67} becomes an estimate for μ_{65} μ_{66} μ_{67} if we make the simplifying assumption that force of mortality is constant in the year of age which can then be used in the calculation for ${}_3p_{65}$

[note for full marks here need to define occurrence exposure rate, link that to m, state assumption under which m = μ and apply to the probability needed]

(e) time selection and spurious selection

IFoA syllabus section 4.1 especially 4.1.3 and 4.1.4

parts a, b, c, e covered in lectures, part d needs number of exercises combined in new (unseen) way

Q2 Kaplan Meier

(a) after j days let

d_j = the number of athletes back running

c_j = the number of athletes who leave the trial (right censoring)

n_j = the risk set of athletes still in rehab

λ_j = the hazard of being back running where $\lambda_j = d_j / n_j$

then for the drug group we have:

j	n	d	c	λ_j	1 - λ_j	S(j)
1	10					1
2	10	1	1	1/10	9/10	0.9
3	8			0	1	0.9
4	8	3		3/8	5/8	0.5625
5	5	3		3/5	2/5	0.225
6	2		2	0	1	0.225
7	0					

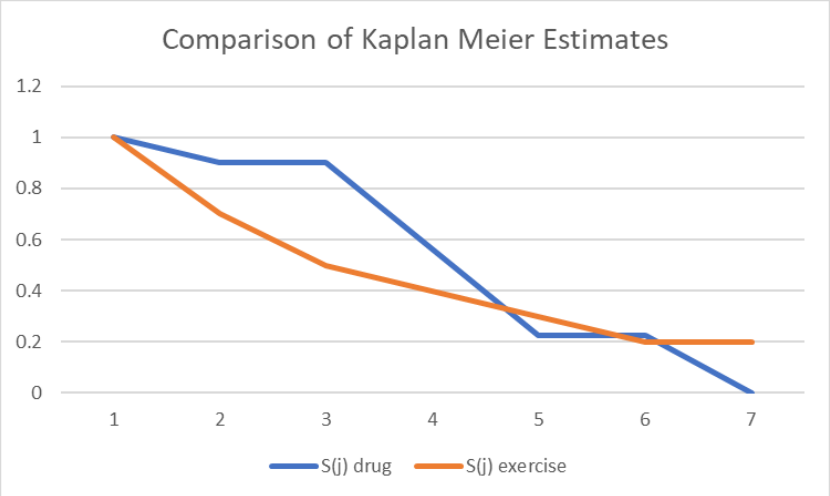
and for the exercise group we have

j	n	d	c	λ_j	1 - λ_j	S(j)
1	10					1
2	10	3		3/10	7/10	0.7
3	7	2		2/7	5/7	0.5
4	5	1		1/5	4/5	0.4
5	4	1		¼	¾	0.3
6	3	1		1/3	2/3	0.2
7	2		2	0	1	0.2

in each case the last column represents the Kaplan Meier estimate of the survival function (the probability of still being in rehab at j days).

(b) both trial groups have right censoring with non-random type II censoring at the end of the drug trial and non-random type I censoring at the end of the exercise trial.

(c) begin by plotting the 2 survival functions



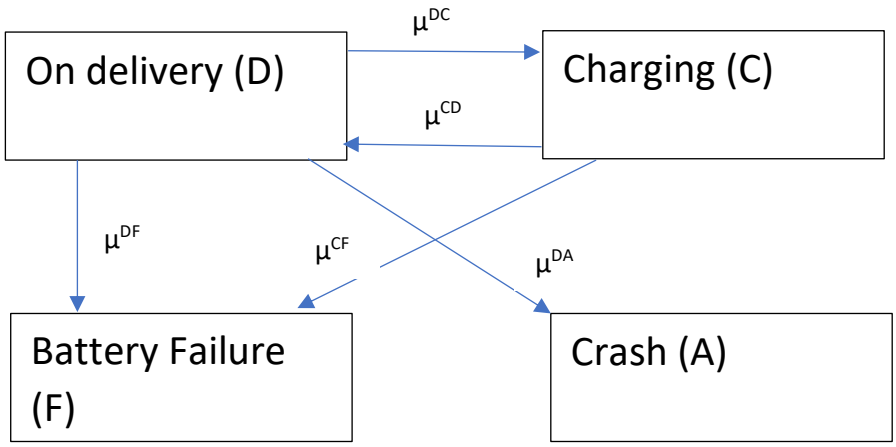
- we seek to minimise the survival function here
- both treatments are effective within a week
- for j<5 it appears that the exercise routine is more effective
- beyond j=5 the situation is less clear
- in particular the effect of the right censoring at j=6,7 for these options needs further investigation, the censoring criteria should really be the same across the two groups
- we would like a larger sample size to draw firmer conclusions

IFoA syllabus section 4.2.1; 4.2.2; 4.2.3

part a similar to seminar question, b covered in lectures, c demands higher level skills

Q3 Multi State model

(a) A multi state model for this scenario is



where the states are D, C, F and A and μ^{PQ} represents the transition intensity between states P and Q

(b) We seek an estimate for μ^{DF} using the notation from (a) above

If the waiting times in states D and C are given by V and W

The number of transitions from D to C is r, from C to D is s, from D to F is t, from D to A is u and from C to F is z then the Likelihood function for this model is given by :

$L(\mu^{DC}\mu^{CD}\mu^{DF}\mu^{DA}\mu^{CF}) = \exp(-(\mu^{DC} + \mu^{DF} + \mu^{D})V).\exp(-(\mu^{CD} + \mu^{CF})W).(\mu^{DC})^{r}(\mu^{CD})^{s}(\mu^{DF})^{t}(\mu^{DA})^{u}(\mu^{CF})^{z}$

We then find log(L) and differentiate wrt μ^{DC} to give the maximum likelihood estimate of μ^{DC} = t/V = 3/1136 = 0.002641

[note for full marks need to state the likelihood function in terms of the notation developed and explain how the likelihood is used to derive MLEs of the transition intensities before moving to the calculation]

(c) the probability of battery failure up to time t is a function of the transition intensities. The multi state model does represent the underlying process exactly and should allow for a consistent, unbiased estimate of the relevant intensities. We note that the number of transitions is very small so any change in r,s,t,u,z would lead to a big change in estimates. This study does not consider other covariates, notably battery age and physical conditions.

IFoA syllabus section 3.3.3 – 3.3.8

part a, b similar to exercises, c unseen and demands higher level skills

Q4 Exposed to Risk

(a) the rate of texts per viewing hour = number of texts / exposed to risk

= ½ x 1286 / E

We cannot calculate E precisely so need to use a Census approximation

We need first to calculate the end time. Run time is 34+35+26+36+3(5) = 146 mins = 2 hours 26 mins so the stream ends at 9:26

Time band	Number hours	Census start	Census end	E-to-Risk
7.00 – 7.30	0.5	214353	256743	117,774.00
7.30 – 8.00	0.5	256743	269459	131,550.50
8.00 – 8.30	0.5	269459	238850	127,077.25
8.30 – 9.00	0.5	238850	234653	118,375.75
9.00 – 9.26	26/60	234653	202464	94,708.68
				589,486.18

where in final column E-to-Risk = ½ x no hours x census start x census end

total Exposed to Risk = 589,486.18 hours

rate of texting = ½ x 1286 / 589486.18 = **0.00109**

(b) The Principle of Correspondence states that a person online at time t should be included in the exposure at time t if and only if, were that life to text immediately, they would be counted in the texts data at that time.

[note do not give marks if principle is stated in terms of mortality rather than the scenario here]

IFoA syllabus section 4.4

part a similar to tutorial example, b in lecture

Q5 Graduation and Statistical Tests

(a) graduation by reference to a parametric formula using Gompertz

advantages

- fits well with probabilistic models of survival
- guaranteed to give smooth function
- estimating parameters is quite straightforward

disadvantages

- we have very large age range here so unlikely one formula will suit all ages
- in particular Gompertz better suited to older ages not younger
- heterogeneity in the data set is an issue here

(b) These tests show:

- an extensive set of tests that allows for different shapes of survival functions in different age ranges
- the Chi-sq tests suggest good overall fit in 3 of 4 age groups but concern in 46-65 group
- the other tests can do 2 things – indicate what the issue in the 46-65 group might be and highlight weaknesses in the other groups that chi-sq fails to pick up
- In 46-65 group the failure of the signs test but not the others suggests that the graduation is consistently under or overestimating mortality in the group
- we cannot tell which from the table
- We note that this age range is where mortality rates start to accelerate from very low levels to the exponential form in older ages
- this suggests a different graduation approach is needed here

- Failure of both signs and grouping of signs at oldest ages would also suggest a over or underestimation here. Is this same direction as middle aged group?
- the standardised devations test in youngest ages suggests some outliers in the data
- This could be due to very small numbers of deaths
- It would also make sense to test the graduation separately for male and female mortality

(c) What adjustments should be made will depend on:

- what the tables are to be used for. In particular note the fit is good in ages 19-45 when most life assurance products are purchased but not in older ages which are more relevant for pensions products
- A Makeham rather than Gompertz parametric formula might work better in middle ages
- Different formulae for different age groups chained together will be more effective still
- comparing chi-sq statistics for different graduation approaches will indicate extent of any improvement
- there is nothing to suggest the underlying Poisson model is the issue here

IFoA syllabus section 4.5.1, 4.5.2, 4.5.5, 4.5.7

part a from lecture, b and c are unseen in this form but use material covered in other exercises and are more challenging