SRM ASSIGNMENT 1

Q1]

$_{\mu}q_x = P[T_x \le \mu]$

$= P[K_x = 0 \;\&\; S_x \le \mu]$

$= P[K_x = 0] * P[S_x \le \mu]$

Since Kx and Sx are independent

$P[S_x \le \mu] = \int_0^{\mu} 1dx = \mu\,, since\ uniform\ distribution$

$P[K_x = 0] = q_x$

$Thus\ _{\mu}q_x = \mu * q_x$

Q2]

$a)\ Central\ exposed\ to\ risk:$

$Period\ of\ exposure\ is\ 1.6.2000\ to\ 25.10.2000$

$= 30 + 31 + 31 + 30 + 25 = 147\ days$

$= \frac{147}{7} = 21\ weeks$

$b)\ Initial\ exposed\ to\ risk$

$Period\ of\ exposure\ is\ 1.6.2000\ to\ 31.5.2000 = 52\ weeks$

Q3]

Left Censoring:

Data in this study would be left censored if the censoring mechanism prevent us

from knowing when the policyholder joined the company.

This is not present because the policy issue date is given.

Right Censoring:

Data would be right censored if the censoring mechanism cuts short observations in

progress, so that we are not able to discover if and when the policy is surrendered.

Data in this study would be right censored if the policy is terminated before the

maturity date for reasons than surrender.

Interval Censoring:

Data in this study would be interval censored if the observational plan only allows us

to say that the duration of policy at the time of surrender fell within some interval of

time.

Here we know the calendar year of surrender and the policy issue date, so we will know that the duration of the policy falls within one year rate interval. Interval

censoring is present.

Informative Censoring:

Censoring in this study would be informative if the censoring event divided individuals into two groups whose subsequent experience was thought to be different.

Here the censoring event of surrendering the policy might be suspected to be informative, as those who are likely to surrender the policy to be in better health

than those who do not surrender the policy.

Q4]

$i)\ Complete\ Expectation\ of\ Life - \dot{e}_x$

$$\dot{e}_x = E[T_x] = \int_0^{\omega - x} {}_tp_x\ dt$$

$This\ represents\ the\ integral\ of\ the\ probability\ of\ survival\ at\ each\ future\ age\ ,$

$in\ short\ it\ means\ the\ ecpected\ future\ lifetime\ of\ a\ person\ aged\ x\ .$

$ii)\ Curtate\ Expectation\ of\ Life - e_x$

$$e_x = \sum_{k=1}^{\infty} {}_kp_0 = \sum_{k=1}^{\infty} e^{-0.0325k} = \frac{e^{-0.0325}}{(1 - e^{-0.0325})} = 30.2719$$

$iii)$ *The probability that a life aged exactly* 36 *will survive to age* 45

$$\rightarrow \quad {}_9p_{36}$$

$${}_9p_{36} = e^{-\int_0^9 0.0325\, dt} = e^{-0.02925} = 0.7464 \ = 74.6\% \ \sim 75\%$$

$iv)$ *The the exact age x representing the median of the life time T of a new born baby*

${}_xp_0 \ = 0.5$, *since the median of lifetime T implies that the prob. of* ${}_xp_0 \ = 0.5$

${}_xp_0 \ = e^{-\mu x}, \qquad \therefore \quad {}_xp_0 \ = 0.5 \ = e^{-0.0325x}$

$0.5 \ = e^{-0.0325x}$

$Taking \log \rightarrow \ \ ln0.5 = \ -0.0325x$

$$\frac{ln0.5}{-0.0325} = x$$

$x = 21.3276 \sim 21.33$

Q5]

i) Gompertz Law is a suitable model for human mortality for middle to older ages say 35 a

There is evidence that the Gompertz Law breaks down at very advanced ages and therefor

is acceptable.

$ii)$ *we know* that ${}_tp_x \ = e^{-\int_0^t \mu_{x+s} ds}$

$$Putting \ \mu_x = Bc^x$$

${}_tp_x = e^{-\int_0^t Bc^{x+s}\ ds}, \qquad c^{x+s} \ = \ c^x \, e^{slogc}$

$$\int_0^t Bc^{x+s} \ \ ds = \int_0^t Bc^x \ e^{slogc} \ \ ds = \frac{Bc^x}{\log c} \ \left[e^{slogc}\right]_0^t$$

$$\frac{Bc^x}{\log c}\,[e^{s\log c}]\,_0^t = \frac{Bc^x}{\log c}\,[c^s]\,_0^t = \frac{Bc^x}{\log c}\,[c^t - 1]\,_0^t$$

$$\therefore\ {}_tp_x = e^{(\log g c^x[c^t-1]} = e^{(\log g)^{c^x(c^t-1)}}$$

$$since\ we\ know\ that\ g\ is\ defined\ as \log g = -\frac{B}{\log c}$$

Q6]

$i)\ The\ hazard\ function\ applies\ to\ Female\ Smoker\ aged\ 30\ at\ entry$

$$ii)\frac{h_{j(t)}}{h_{i(t)}} = \frac{e^{-0.05}}{e^{0.1}} = 0.86070$$

$here, j\ is\ the\ male\ smoker\ aged\ 30\ at\ entry\ and\ i\ is\ the\ female\ smoker\ aged\ 40$

$we\ know, S(t) = e^{-\int_0^t h(s)ds}$

$\therefore s_j\ (t) = (s_i(t)\ )^{0.86070}$

$this\ implies\ that\ \ s_j\ (t) > s_i(t)\ for\ all\ t > 0$

$$iii)\ \frac{h_{j(t)}}{h_{i(t)}} = \frac{e^{0.2}}{e^{0.05}} = 1.161$$

$Here, j\ is\ the\ male\ smoker\ aged\ 30\ at\ entry\ and\ i\ is\ the\ male\ smoker\ aged\ 40\ at\ entry$

$we\ know, S(t) = e^{-\int_0^t h(s)ds}$

$s_j\ (t) = (s_i(t)\ )^{1.161}$

$this\ implies\ that\ \ s_j\ (t) > s_i(t)\ for\ all\ t > 0$

Q7]

i) The most appropriate rate interval to use (for lives classified x) is the policy year rate interval starting on the policy anniversary where lives are aged x next birthday.

The reason is that this corresponds to the definition of the deaths and the rate is more sensitive to errors in approximation of the numerator than the denominator.

The average age at the start of the rate interval is $x - \frac{1}{2}$ assuming that birthdays are uniformly distributed over the policy year.

ii) We will use the following symbols :

$P_{x,t}$: to represent the in force at time t from the 1 January 1997 classified x next

birthday on policy anniversary nearest to time t

$\theta_{x,t}$: to represent the deaths in the calendar year 1997 aged x next birthday on policy

anniversary (= age next birthday at entry plus curtate duration at date of death)

before death

E_x, E_x^c : to represent the initial and central exposed to risk respectively of lives age x

last birthday on previous policy anniversary.

$P_{x(t)}$: to represent the in force at time t from the 1 January 1997 classified x next

birthday on the policy anniversary preceding time t.

$$Now\ P_x(t) = \frac{1}{2}\left(P_{x,t} + P_{x+1,t}\right)$$

$$assuming\ that\ policy\ anniversaries\ are\ uniformly\ distributed$$

$over\ the\ calendar\ year.$

$$E_x^c = \int_0^{10} P_x(t)dt = \frac{1}{2}\sum_{t=0}^{9}[P_x(t) + P_x(t+1)]$$

$inforce\ population\ varies\ linearly\ betwn\ the\ dates\ of\ investigation$

$$E_x = E_x^c + \frac{1}{2}\sum_{t=0}^{10}\theta_{x,t}$$

$assuming\ that\ in\ aggregate\ the\ deaths\ occur\ on\ average\ halfway$

$through\ the\ policy\ year$

Q8]

i) Types of censoring presents:

1. Type I censoring present because the study ends at a predetermined duration of 45 days.
2. Type II censoring is not present because the study did not end after a predetermined number of patients had died.
3. Random censoring is present because the duration at which a patient left
   hospital before the study ended can be considered as a random variable.
4. Right Censoring is present for those lives that exit before the end of investigation period

ii) The censoring is likely to be informative.

The patients who died were probably recovering less well that patient who discharged from the hospital.

If they had not died, they would likely to remain in the hospital for longer than those who were not censored.

iii) The Kaplan-Meier estimate of the survival function is estimated as follows:

t_j	n_j	d_j	c_j	λ_j	$1-\lambda_j$	S(t)
0	13	0				
5	13	1	0	0.0769	0.9231	0.92
7	12	1	0	0.0833	0.9167	0.85
14	11	1	2	0.0909	0.9091	0.77
28	8	1	2	0.1250	0.8750	0.67
35	5	1		0.2	0.8	0.54

So, the value survival function at end of investigation period is 0.54

Assumptions:

- The censoring happens just after the death

- Ignoring the discharge on any other ground except recovery from illness

- Ignore any admission period before the start of investigation [4]

iv) Comments:

1. The survival of a patient from the infection who given treatment is around 50% in light
   of the answer in c_j above.
2. However, the hospital excluded the number of deaths who died within two weeks of
   observation period.
3. It also ignores the admission pre investigation period
4. It is assuming that the censored patient at the end of investigation will survive for sure.
5. Also ignoring the patients being discharged on any other ground like shifting to

another hospital etc.

6. It claims that 8 out of 10 patients who responded the treatment beyond two weeks
would survive.
7. So, the claims have to be viewed with respect to above considerations.

Q9]

$$a)\ Under\ the\ UDD\ assumption$$

$$\int_0^1 tp_x\, dt = \int_0^1 (1 - tq_x)dt = [t - 0.5t^2\, q_x\,]_0^1 = 1 - 0.5q_x$$

$$q_x = 0.3\,, we\ have$$

$$m_x = \frac{0.3}{1 - 0.15}$$

$b)Under\ CFM\ assumption$

$$\int_0^1 tp_x\, dt = \int_0^1 e^{-\mu t}\, dt = \frac{1}{\mu}(1 - e^{-\mu}) = q_{\frac{x}{\mu}}$$

$$so, m_x = \mu = -\ln(1 - q_x) = 0.356675$$

Q10]

i) Under the Cox model each individual's hazard is proportional to the baseline hazard, with the constant

of proportionality depending on certain measurable quantities called co-variates. Hence the model is

also called a proportional hazards model.

Q11]

i) Consider the durations tj at which events take place.

Let the number of deaths at duration tj be dj and the number of insects still at risk of death at duration

tj be nj.

At tj = 1, S(t) falls from 1.0000 to 0.9167.

$S(t) = \prod_{t_j \le t}(1 - \lambda(t_j)) = 0.9167(1\text{-}\lambda(3))$

$we\ must\ have\ 0.9167 = 1 - \lambda(1), so\ that\ \lambda\ (1)\ is\ 0.0833$

$since\ \lambda(1) = \frac{d_1}{n_1}\ , then\ we\ have\ \ \frac{d_1}{n_1} = 0.0833$

and, since all 12 insects are at risk of dying at tj = 1, we must therefore have d1 = 1 and n1 = 12.

Similarly, at tj = 3, we must have $0.7130 = 0.9167\big(1 - \lambda(3)\big)$

$$\lambda\ (3)\ = \frac{0.9167 - 0.7130}{0.9167} = 0.222\ = \frac{d_3}{n_3}$$

Since we can have at most 11 insects in the risk set at tj = 3, we must have d3 = 2 and n3 = 9.

Similarly, at tj = 6, we must have $0.4278 = 0.7130\big(1 - \lambda(6)\big)$

$$\lambda\ (6)\ = \frac{0.7130 - 0.4278}{0.7130} = 0.4\ = \frac{d_6}{n_6}$$

Since we can have at most 7 insects in the risk set at tj = 6, we must have d6 = 2 and n6 = 5.

Therefore 2 insects died at duration 3 weeks and 2 insects died at duration 6 weeks.

t	n	d	c	$\lambda(t)$	S(t)
0	12	0		0	1.000
1	12	1	2	0.833	0.9167
3	9	2	2	0.22	0.7130
6	5	2	3	0.4	0.4278

iii) Summing up the number of deaths we have total deaths = d1+d3+d6= 1+2+2= 5.

Since we started with 12 insects, the remaining 7 insects' histories were right censored.

Q12]

i) Gompertz Law:

Gompertz Law is an exponential function, and it is often a reasonable assumption for middle

and older ages. It can be expressed as follows:

$\lambda_x = Bc^x \; ; where \; \lambda_x \; is \; a \; force \; of \; martality \; at \; age \; x$

$ii) \; Substituting \;\; B = e^{\beta_0+\beta_1 X_1+\beta_2 X_2} \; ; into \; gompertz \; model$

$\lambda_x = e^{\beta_0+\beta_1 X_1+\beta_2 X_2} \; * c^x \; ; defining \; x \; as \; a \; duration \; since \; 50th \; b'day$

$The \; hazard \; can \; therefore \; be \; factorized \; into \; 2 \; parts$

$e^{\beta_0+\beta_1 X_1+\beta_2 X_2} \; , which \; depends \; only \; on \; the \; valus \; of \; the \; covariates \; and \; c^x,$

$which \; depends \; only \; on \; duration.$

$so, the \; ratio \; betwn \; the \; hazards \; for \; any \; 2 \; persons \; with \; different$

$characteristics \; does \; not \; depend \; on \; duration \; , so \; the \; model \; is$

$proportional \; hazard \; model.$

$iii) \; The \; baseline \; hazard \; in \; this \; model \; relates \; to \; non \; smoker \; female$

$iv) \; For \; the \; female \; cigarette \; smoker$

$X_1 = 0 \;\; and \; X_2 = 1 \; amd \; x = 4$

$Therefore \; the \; hazard \; at \; age \; 54 \; is :$

$\lambda_x = e^{\beta_0+\beta_1.0+\beta_2.1} * c^4$

$= e^{-4+0.65} * 1.05^4$

$= 0.0351 * 1.2155$

$= 0.04266$

$v)\ the\ hazard\ for\ a\ non\ smoker\ at\ duration\ s\ is\ given\ by:$

$\lambda_s = e^{\beta_0+\beta_1.X_1} * c^s$

$the\ hazard\ for\ a\ non\ smoker\ at\ duration\ t\ is\ given\ by:$

$\lambda_t * = e^{\beta_0+\beta_1.X_1+0.65} * c^t$

$if\ the\ smoker'sand\ non\ smoker'shazards\ are\ same\ ,$

$then\ \lambda_s = \ \lambda_t *$

$i.e.\ e^{\beta_0+\beta_1.X_1} * c^s = e^{\beta_0+\beta_1.X_1+0.65} * c^t$

$i.e.\ c^s = e^{0.65}.c^t$

$i.e.\ c^{s-t} = e^{0.65} = 1.9155$

$since, c = 1.05$

$hence, 1.05^{s-t} = 1.9155$

$$so\ s - t = \frac{\ln(1.9155)}{\ln(1.05)} = 13.32$$

Hence, when the two hazards are equal, the non-smoker is approximately 13 years older than the smoker.

Q13]

i) Let P'x(t) be the number of policies inforce aged x nearest birthday at time t .

Also, let Px(t) be the number of policies inforce aged x last birthday at time t

Let E_x^c refers to the central exposed to risk at age label x respectively.

$$E_x^c = \int_{t=o}^{2} P'x(t)\ dt$$

$assuming\ that\ P'_{56}(t) is\ linear\ over\ the\ year$ (2015,2016) and (2016,2017)

$we\ can\ approximate\ the\ exposure\ as\ follows$

$$E_x^c \frac{1}{2} * P'_{56}(2015) + P'_{56}(2016) + \frac{1}{2} P'_{56}(2016) + P'_{56}(2017)$$

$$= \frac{1}{2} P'_{56}(2015) + P'_{56}(2016) + \frac{1}{2} P'_{56}(2017)$$

Since, the number of policyholders aged label 56 nearest birthday will be between 55.5 and

56.5 i.e. between age label 55 last birthday and 56 last birthday. Assuming that the

birthdays are uniformly distributed over the calendar year:

$$P'_{56}(2015) = \frac{1}{2}\left(P'_{55}(2015) + P'_{56}(2015)\right) = 20050$$

$$P'_{56}(2016) = \frac{1}{2}(P'_{55}(2016) + P'_{56}(2016) = 20800$$

$$P'_{56}(2017) = \frac{1}{2}(P'_{55}(2017) + P'_{56}(2017) = 19250$$

$$E_{56}^c = \frac{1}{2} * 20050 + 20800 + \frac{1}{2} * 19250 = 40450$$

$$\mu_{56} = \frac{d_{56}}{E_{96}^c} = 0.0341$$

$deriving\ the\ force\ of\ mortality\ for\ age$ 57 $as\ above$

$$P'_{57}(2015) = \frac{1}{2}\left(P'_{56}(2015) + P'_{57}(2015)\right) = 19850$$

$$P'_{57}(2016) = \frac{1}{2}\left(P'_{56}(2016) + P'_{57}(2016)\right) = 20900$$

$$P'_{57}(2017) = \frac{1}{2}\left(P'_{56}(2017) + P'_{57}(2017)\right) = 17500$$

$E_{57}^{c} = \frac{1}{2} * 19850 + 20900 + \frac{1}{2} * 17500 = 39575$

$\mu_{57} = 0.03588$

$ii)\ q_{55.5} = 1 - e^{\mu_{56}} = 0.0335$

$q_{56.5} = 1 - e^{\mu_{57}} = 0.0352$

Q14]

i) The hazard function for getting married is given by:

(t,Z) = 0(t)exp[0.3 Z1 + 0.2 Z2 + 0.3 Z3 + 0.5 Z4 – 0.1 Z5 + 0.7 Z6 + 0.5 Z7 – 0.4 Z8]

Where

0(t) = baseline hazard at time t since looking for the life partner.

Z = (Z1, Z2, Z3, Z4, Z5, Z6, Z7, Z8)

Z1 = 1 if female, 0 if not.

Z2 = 1 if location = Non Metro, 0 if not

Z3 = 1 if profession = Service, 0 if not

Z4 = 1 if profession = Business, 0 if not

Z5 = 1 if profession = Social Service, 0 if not

Z6 = 1 if Age Band = 20-25, 0 if not

Z7 = 1 if Age Band = 25-30, 0 if not.

Z8 = 1 if Age Band = 35-40, 0 if not. [2]

ii) People most likely to stay single with the lowest hazard function.

The probability that a person who has been looking for a life partner for one year will stay single for

next 2 years is:

$e^{-\int_1^3 \lambda(t,z)dt}$

If the person is a female, profession as a social service and aged 37, the probability is:

$P_F = \exp\left[-e^{0.3z}1 \quad {}^{-0.1z}_{5} \, {}^{-0.4z}_{8} \quad integral \left|{}_1^3 \lambda_0(t)dt\right]\right.$

Let A $=$ e^ $\int_1^3 \lambda_0(t)dt$

$P_F = 0.3$

$A = 0.2298$

Q15]

i. Advantages of central exposed to risk.

Two advantages of central exposed to risk over initial exposed to risk are:

1. The central exposed to risk is simpler to calculate from the data typically available compared to the initial exposed to risk. Moreover, central exposed to risk has an intuitive appeal as the total observed waiting time and is easier to understand than the initial exposed to risk.

2. It is difficult to interpret initial exposed to risk in terms of the underlying process being modelled if the number of decrements under study increase or the situations become more elaborate. On the contrary, the central exposed to risk is more versatile and it is easy to extend the concept of central exposed to risk to cover more elaborate situations.

ii. Calculation of exposed to risk.

Rita

Rita turned 30 on 1 October 2009, when she was already married. She died on 1 January 2010, 3 months after her 30th birthday.

Thus, Rita‟s contribution to central exposed to risk = 3 months

And contribution to initial exposed to risk = 1 year

Sita

Sitaturned 30 on 1 September 2011, when she was already married. Time spent under investigation, aged 30 last birthday by Sita was 1 September 2011 – 31 August 2012.

Thus, Sita‟s contribution to both central and initial exposed to risk is 1 year.

Nita

Nita turned 30 on 1 December 2009 and married 2 months later. Therefore, she joined the investigation of married women on 1 February 2010. She divorced 9 months later, when she would be censored from the investigation of married women.

Thus, Nita‟s contribution to both central and initial exposed to risk is 9 months.

Gita

Gita got married on 1 June 2011, at which time she was already past her 31st birthday. Therefore, she has spent no time during the investigation period as a married woman at age 30 last birthday.

Thus, her contribution to both central and initial exposed to risk is nil.

iii. Total exposed to risk.

Hence, total exposed to risk is:

Central exposed to risk $= 0.25 + 1 + 0.75 + 0 = 2$ years.

Initial exposed to risk $= 1 + 1 + 0.75 + 0 = 2.75$ years

From the results above, it can be seen that the central exposed to risk is 2 years and the initial exposed to risk is 2.75 years. The approximation would suggest that the initial exposed to risk should be 2.5 years. However, this is not a good approximation for the data provided as the approximation is based on the assumption that deaths would be evenly spread and thus can be assumed to occur half way through the year, on average. This also relies on an implicit assumption of a reasonably large data set. In the data above, there were only 4 lives, which is not statistically significant. Moreover, there was only one death, which occurred 3 months after the 30thbirthday. As a result of the statistical sparseness in the data, the approximation is seen not to work very well.