Statistical and Risk Modelling - Assignment 1
Palak Nagdev, Roll number 22

1. $P(T_x \leq u) = \ _uq_x$

$\therefore P(K_x = 0 \ \ and \ \ S_x \leq u) = \ _uq_x$ ... $(0 < u < 1 \ and \ T_x = K_x + S_x)$

Because K_x and S_x are independent,

$\therefore P(K_x = 0\ ) * P(\ S_x \leq u) = \ _uq_x$

$\therefore \ q_x * u = \ _uq_x$

$\therefore \ _uq_x = u \cdot q_x \quad where\ 0 < u < 1$

Hence Proved.

2.

A-

The life will be age 54 last birthday between 1/6/2000 and 31/5/2001.

But the life died on 25/10/2000

∴ The central exposed to risk of the life aged 54 last birthday is from 1/6/2000 to 24/10/2000. (Including the first day and excluding the last day)

number of days = 146 days

number of weeks = 146/7

= 20.85714 weeks

$\approx 21\ weeks$

B-

The life was alive on 1/6/2000 when he attained the exact age 54

Therefore, Initial exposed to risk of the life aged 54 last birthday is from 1/6/2000 to 31/5/2001.

= 52 weeks

3.]

<u>Left censoring</u>

Left censoring occurs when a data point is below a certain value but is unknown by how much.

Left censoring is not present in the dataset

<u>Right censoring</u>

Right censoring occurs when a data point is above a certain value but is unknown by how much.

Right censoring is present in the dataset. If any life does not surrender his/her policy then the life is right censored. Also, if the exit from policy is due to any other reasons, then the policy is not surrendered and hence is an example of right censoring. If the policy matures instead of being surrendered then it is an example of right censoring as well.

<u>Interval Censoring</u>

Interval censoring occurs when a data point is somewhere in an interval between two periods.

Interval censoring is present in the dataset as we do not know the exact date of surrender but only the calendar year.

<u>Informative censoring</u>

When censoring of a data point occurs, if we get additional information from the censoring then it can be termed as informative censoring.

Informative censoring is present in the dataset as the policy holder can exit due to other reasons and hence, we can get certain additional information about the life.

4. (i)

Complete expectation of life ($\dot{e}_x$) = $\int_0^{\omega-x} tPx\ dt$ = $\int_0^{\omega-x} S_x(t)\ dt$

E[T_x] is the expected complete future lifetime of a life aged 'x'.

(ii) The curate expectation of life

$$tPx = e^{-\int_0^t u_{x+s} ds}$$

For constant force of mortality,

$$tPx = e^{-u*t}$$

E[K_x] = e_x = $\sum_{k=1}^{[\omega-x]} {}_k p_x$

$$e_x = \sum_{k=1}^{[\omega-x]} e^{-u*k}$$

$$e_x = e^{-u} + e^{-2u} + e^{-3u} + \cdots + e^{-[\omega-x]*u}$$

$$e_x = \frac{e^{-u}}{1-e^{-u}}$$

$$e_x = \frac{e^{-0.0325}}{1-e^{-0.0325}}$$

$$e_x = 30.27194\ years$$

(iii) Probability that a life aged exactly 36 will survive to age 45

For constant force of mortality,

$$tPx = e^{-u*t}$$

$$9P36 = e^{-0.0325*9}$$

$$9P36 = 0.746395245$$

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

$$P(T_0 \le x) = 0.5$$

$$xq0 = 0.5$$

$$1 - xP0 = 0.5$$

$$1 - e^{-u*x} = 0.5$$

$$1 - e^{-0.0325*x} = 0.5$$

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

$$\ln(0.5) = -0.0325 * x$$

$$\therefore \mathrm{x} = 21.327 \text{ years}$$

5.

(i)

Gompertz law is an appropriate model for human mortality for range 30 to 80 years.

(ii)

For Gompertz law

$$u_x = Bc^x$$

We also know,

$$u_{x+t} = -\frac{1}{tPx} * \frac{d}{dt} tPx$$

$$u_{x+t} = -\frac{d}{dt} \ln{(tPx)}$$

$$Bc^{x+t} = -\frac{d}{dt} \ln{(tPx)}$$

Integrating on both sides,

$$\int_0^t Bc^{x+s}\, ds = -\ln(tPx)$$

$$\int_0^t Bc^{x+s}\, ds = -\ln(tPx)$$

$$\int_0^t Bc^x c^s\, ds = -\ln(tPx)$$

$$\int_0^t Bc^x c^s\, ds = -\ln(tPx)$$

$$Bc^x \int_0^t c^s\, ds = -\ln(tPx)$$

$$Bc^x \int_0^t e^{lnc^s}\, ds = -\ln(tPx)$$

$$Bc^x \int_0^t e^{lnc^s}\, ds = -\ln(tPx)$$

$$Bc^x \int_0^t e^{slnc}\, ds = -\ln(tPx)$$

$$\frac{Bc^x}{lnc} [e^{lnc^s}]_0^t = -\ln(tPx)$$

$$\frac{Bc^x}{lnc} [c^s]_0^t = -\ln(tPx)$$

$$\frac{Bc^x}{lnc}[c^t - c^0] = -\ln(tPx)$$

$$\frac{Bc^x}{lnc}[c^t - 1] = -\ln(tPx)$$

$$\frac{Bc^x}{lnc}[c^t - 1] = -\ln(tPx)$$

$$-[c^t - 1] = \ln(tPx)$$

$$e^{-\frac{Bc^x}{lnc}[c^t-1]} = tPx$$

$$e^{-\frac{B}{lnc}c^x[c^t-1]} = tPx$$

Substituting $lng = -\frac{B}{lnc}$

$$e^{(lng)\,c^x[c^t-1]} = tPx$$

$$g^{c^x[c^t-1]} = tPx$$

Hence Proved.

Under Gompertz Law,

$$tPx = g^{c^x[c^t-1]} \quad where\ g = e^{-\frac{B}{lnc}}$$

6. The study is been conducted to investigate the effect of a newly invented drug on a group of patients who are suffering from cancer.

The hazard function for life 'i' at duration 't', ki given by

$$h_i(t) = h_0(t) * e^{\{0.01*(x_i - 30) + 0.2*y_i - 0.05*z_i}$$

Where,

$$h_i(t) \text{ denotes the hazard function for life } i \text{ at duration } t$$

$$h_0(t) \text{ denotes the baseline haxard function at duration } t$$

$$x_i \text{ denotes the age of entry into the observation of life } i$$

$$y_i = 1 \text{ if life } i \text{ a non} - \text{smoker, else } 0$$

$$z_i = 1 \text{ if life } i \text{ a male, } 0 \text{ if female}$$

(a)

The baseline hazard function applies to a female smoker whose age is 30 years at the time of entry into the observation.

(b)

To compare the survival function of a male smoker aged 30 at entry relative to a female smoker aged 40 at entry

For a male smoker aged 30 at entry,

$$h_{male,smoker,aged\ 30}(t) = h_0(t) * e^{-0.05} = h_0(t) * 0.9512294$$

For a female smoker aged 40 at entry,

$$h_{female,smoker,aged\ 40}(t) = h_0(t) * e^{0.1} = h_0(t) * 1.105171$$

$$\frac{h_{male,smoker,aged\ 30}(t)}{h_{female,smoker,aged\ 40}(t)} = \frac{h_0(t) * 0.9512294}{h_0(t) * 1.105171} = 0.860708$$

$$S_x(t) = e^{-\int_0^t h_i(s) ds}$$

For male smoker aged 30 at entry,

$$S_{male,smoker.aged\ 30}(t) = e^{-\int_0^t h_{male,smoker,aged\ 30}(s)\ ds}$$

$$S_{male,smoker.aged\ 30}(t) = e^{-\int_0^t 0.860708*h_{female,smoker,aged\ 40}(s)\ ds}$$

$$S_{male,smoker.aged\ 30}(t) = e^{-0.860708\int_0^t h_{female,smoker,aged\ 40}(s)\ ds}$$

$$S_{male,smoker.aged\ 30}(t) = (e^{-\int_0^t h_{female,smoker,aged\ 40}(s)\ ds})^{0.860708}$$

$$S_{male,smoker.aged\ 30}(t) = (S_{female,smoker.aged\ 40}(t))^{0.860708}$$

Therefore,

$$S_{male,smoker.aged\ 30}(t) > S_{female,smoker.aged\ 40}(t) \quad (for\ all\ t)$$

(c)

To compare the survival function of a male smoker aged 30 at entry relative to a male smoker aged 40 at entry

For a male smoker aged 30 at entry,

$$h_{male,smoker,aged\ 30}(t) = h_0(t) * e^{-0.05} = h_0(t) * 0.9512294$$

For a male smoker aged 40 at entry,

$$h_{male,smoker,aged\ 40}(t) = h_0(t) * e^{0.05} = h_0(t) * 1.0512711$$

$$\frac{h_{male,smoker,aged\ 30}(t)}{h_{male,smoker,aged\ 40}(t)} = \frac{h_0(t) * 0.9512294}{h_0(t) * 1.0512711} = 0.9048374$$

$$S_x(t) = e^{-\int_0^t h_i(s)ds}$$

For male smoker aged 30 at entry,

$$S_{male,smoker.aged\ 30}(t) = e^{-\int_0^t h_{male,smoker,aged\ 30}(s)\ ds}$$

$$S_{male,smoker.aged\ 30}(t) = e^{-\int_0^t 0.9048374 * h_{male,smoker,aged\ 40}(s)\ ds}$$

$$S_{male,smoker.aged\ 30}(t) = e^{-0.9048374\int_0^t h_{male,smoker,aged\ 40}(s)\ ds}$$

$$S_{male,smoker.aged\ 30}(t) = (e^{-\int_0^t h_{male,smoker,aged\ 40}(s)\ ds})^{0.9048374}$$

$$S_{male,smoker.aged\ 30}(t) = (S_{male,smoker.aged\ 40}(t))^{0.9048371}$$

Hence

$$S_{male,smoker.aged\ 30}(t) > S_{male,smoker.aged\ 40}(t) \quad (for\ all\ t)$$

7. (i)

The rate interval that would be used for the investigation would be age nearest birthday.

The reason for this, is that the number of deaths is classified in terms of age next birthday.

(ii)

Let $P_{x,t} = Number\ of\ policies\ in\ force\ classified\ age\ next\ birthday$

$$E_x^c = \int_0^{10} P_x(t)\,dt$$

$$E_x^c = \sum_0^9 1 * \frac{1}{2} * P(x,t) + P(x,t+1)$$

$$E_x = E_x^c + \frac{1}{2}\sum_0^{10} dx$$

8.

The types of censoring present in the investigation are as follows:

a) Right censoring

Right censoring is present in the dataset because the decrement in life occurs even when they are discharged or a period of 45 days are elapsed. Since, they didn't die, we were not able to study the mortality of the lives which was the objective of the study.

b) Type I censoring

Sine, there is a fixed time period of 45 days for the investigation, Type I censoring is present.

c) Random censoring

Since, the lives leave the investigation for reasons other than death and the time of exit is not known in advance, random censoring is present in the dataset.

d) Informative censoring

The lives that are discharged can be assumed to have low mortality; hence informative censoring is present in the investigation.

(ii)

Yes, the censoring is likely to be informative in case where the person leaves the investigation owing to discharge.

(iii)

Assuming that the censoring present in the investigation is non-informative

Kaplan-Meier estimate of the survivor function for the patients

j	t_j	d_j	n_j	$\hat{\lambda}_j = \frac{d_j}{n_j}$	1-$\hat{\lambda}_j$	$\prod_{t_j \le t}(1-\hat{\lambda}_j)$
1	5	1	13	0.07692	0.92308	0.92308
2	7	1	12	0.08333	0.91667	0.8461597
3	14	1	11	0.09091	0.90909	0.769235
4	28	1	8	0.125	0.875	0.673081
5	35	1	5	0.2	0.8	0.538465

Range	$\hat{s}(t)$
0 ≤ t < 5	1
5 ≤ t < 7	0.92308
7 ≤ t < 14	0.8461597
14 ≤ t < 28	0.769235
28 ≤ t < 35	0.673081
35 ≤ t < 45	0.538465

(iv)

$\hat{s}(14) = 0.769235.$

Therefore, the hospital claim is not true. Since, the survival probability after 14 days 0.769235.

9.

(i)

(a)

$$m_x = \frac{q_x}{\int_0^1 tPx\ dt}$$

$$m_x = \frac{0.3}{\int_0^1 1 - tqx\ dt}$$

For uniform distribution of deaths between x and x+1

$$m_x = \frac{0.3}{\int_0^1 1 - t * qx\ dt}$$

$$m_x = \frac{0.3}{\int_0^1 1 - 0.3t\ dt}$$

$$m_x = \frac{0.3}{[t - 0.3 * \frac{t^2}{2}]_0^1}$$

$$m_x = \frac{0.3}{[1 - 0.3 * \frac{1}{2}]}$$

$$m_x = 0.3529412$$

(b)

To find: m_x assuming constant force of mortality between ages x and x+1

$$m_x = \frac{q_x}{\int_0^1 tPx\ dt}$$

For constant force of mortality between ages x and x+1

$$tqx = 1 - e^{-u*t}$$

$$1 - \ tqx = \ e^{-u*t}$$

$$1 - \ tqx = \ e^{-u*t}$$

$$1 - \ qx = \ e^{-u}$$

$$e^{-u} = 0.7$$

$$m_x = \frac{0.3}{\int_0^1 e^{-u*t}\ dt}$$

$$m_x = \frac{0.3}{\int_0^1 0.7^t \, dt}$$

$$m_x = \frac{0.3 * \ln{(0.7)}}{[0.7^t]_0^1}$$

$$m_x = \frac{0.3 * \ln{(0.7)}}{[0.7^1 - 1]}$$

$$m_x = 0.35667$$

10.

(i)

Cox model is also called as a proportional hazard model since, the hazard function of a life is proportional to a baseline hazard function.

(ii)

Equation for Cox Proportional Model

$$h_i(t) = \ h_0(t) * e^{\beta_F * F + \ \beta_D * D + \ \beta_M * M}$$

Where,

$h_i(t) = Hazrad\ function\ for\ life\ i\ at\ duration$ t

$h_0(t) = Baseline\ hazrad\ function\ for\ life\ at\ duration$ t

$\beta_F, \beta_D, \beta_M = Parameterd\ for\ F, D\ and\ M$

F = 0 if premium frequency is annual and 1 if non-annual

D = 0 if online, 1 if Agency and 2 if Bancassurance

M = 0 if Direct debit and 1 if cheque

(iii)

The baseline hazard function applies to a policy for which the premium frequency is annual, the distribution channel is online and the method of premium payment is direct debit.

(iv)

$$\frac{h_{annual,agency,direct\ debit}}{h_{non-annual,agency,cheque}} = 0.75 \quad - - - (1)$$

Also,

$$\frac{h_{annual,agency,direct\ debit}}{h_{non-annual,online,direct\ debit}} = 1 \quad - - - (2)$$

Also,

$$h_{annual,online,cheque} = \frac{3}{4} * h_{annual,bancassurance,direct\ debit}$$

$$\frac{h_{annual,online,cheque}}{h_{non-annual,bancassurance,direct\ debit}} = \frac{3}{4} \quad - - - (3)$$

In equation (1)

$$\frac{h_{annual,agency,direct\ debit}}{h_{non-annual,agency,cheque}} = 0.75$$

$$\frac{h_0(t) * e^{\beta_D}}{h_0(t) * e^{\beta_F + \beta_D + \beta_M}} = 0.75$$

$$\frac{e^{\beta_D}}{e^{\beta_F + \beta_D + \beta_M}} = 0.75$$

$$e^{-\beta_F - \beta_M} = 0.75$$

$$\beta_F + \beta_M = 0.287682 \quad ---(4)$$

In equation (2)

$$\frac{h_{annual,agency,direct\ debit}}{h_{non-annual,online,direct\ debit}} = 1$$

$$\frac{h_0(t) * e^{\beta_D}}{h_0(t) * e^{\beta_F}} = 1$$

$$\frac{e^{\beta_D}}{e^{\beta_F}} = 1$$

$$e^{\beta_D - \beta_F} = 1$$

$$\beta_D - \beta_F = 0 \quad ---(5)$$

In equation (3)

$$\frac{h_{annual,online,cheque}}{h_{annual,bancassurance,direct\ debit}} = \frac{3}{4}$$

$$\frac{h_0(t) * e^{\ \beta_M}}{h_0(t) * e^{\ 2\beta_D}} = 0.75$$

$$\frac{e^{\beta_M}}{e^{2\beta_D}} = 0.75$$

$$e^{\beta_M - 2\beta_D} = 0.75$$

$$\beta_M - 2\beta_D = -0.287682 \quad ---(6)$$

From (5)

$$\beta_D = \beta_F ---(7)$$

From (6) and (7)

$$\beta_M - 2\beta_F = -0.287682 \quad ---(8)$$

From (4) and (8)

Subtracting (8) from (4)

$$3\beta_F = 0.575364$$

$$\beta_F = 0.191788 \quad - - - (9)$$

From (7)

$$\beta_D = 0.191788$$

From (4) and (9)

$$\beta_M = 0.095894$$

11.

(i)

Kaplan-Meier estimate of survival function:

t (weeks)	$\hat{s}(t)$
0 ≤ t < 1	1.0000
1 ≤ t < 3	0.9167
3 ≤ t < 6	0.7130
6 ≤ t	0.4278

We know,

$$\hat{s}(t) = \prod_{t_{j} \leq t} (1 - \hat{\lambda} j)$$

For 0 ≤ t < 1

$$\hat{s}(t) = 1$$

$$\left(1 - \frac{d_j}{n_j}\right) = 1$$

$$\frac{d_j}{n_j} = 0$$

$$d_j = 0 \; and \; n_j = 12$$

For 1 ≤ t < 3

$$\hat{s}(t) = 0.9167$$

$$\left(1 - \frac{d_1}{n_1}\right) = 0.9167$$

$$\frac{d_1}{n_1} = 0$$

$$d_1 = 0 \; and \; n_1 = 12$$

For 3 ≤ t < 6

$$\hat{s}(t) = 0.7130$$

$$\left(1 - \frac{d_3}{n_3}\right) = 0.77779$$

$$\frac{d_3}{n_3} = 0.22221$$

Here, if we take $n_3 = 11$, we get $d_3 = 2.44431$

If we take $n_3 = 10$, we get $d_3 = 2.2221$

If we take $n_3 = 9$, we get $d_3 = 2$

d_j can't take non-integer values

$$d_3 = 2 \; and \; n_3 \; = 9$$

For $6 \leq t$

$$\hat{s}(t) = 0.4278$$

$$\left(1 - \frac{d_6}{n_6}\right) = 0.6$$

$$\frac{d_6}{n_6} = 0.4$$

Here, we can have maximum of 7 at risk insects.

If we take $n_6 = 7$, we get $d_6 = 2.8$

If we take $n_6 = 6$, we get $d_6 = 2.4$

If we take $n_6 = 5$, we get $d_6 = 2$

d_j can't take non-integer values

$$d_6 = 2 \; and \; n_6 \; = 5$$

The number of insects dying at duration 3 weeks is 2 and that at 6 weeks is also 2.

(ii)

Sine, we have a total of 12 insects at the start of the investigation and the total number of insects died are 5

Therefore, number of insects whose history is censored = 12 – 5 = 7

12.

(i)

Gompertz law is a defined parametric survival model.

Gompertz function is an exponential function and it is often a reasonable assumption for Middle Ages and older ages. The function increases exponentially with age.

Gompertz law: $\lambda_x = B * c^x$

(ii)

Substituting $B = e^{\beta_0+\beta_1 x_1+\beta_2 x_2}$ in Gompertz law

$$\lambda_x = e^{\beta_0+\beta_1 x_1+\beta_2 x_2} * c^x$$

Here the hazard function of a life is divided in two parts: $e^{\beta_0+\beta_1 x_1+\beta_2 x_2}$ which depends only on the covariates and c^x which depends only on duration.

Thus, the substitution leads to a proportional hazard model

(iii)

The baseline hazard applies to a life who is a female non-smoker

(iv)

$$\lambda_x = e^{\beta_0+\beta_1 x_1+\beta_2 x_2} * c^x$$

$$\lambda_4 = e^{-4+0.65} * (1.05)^4$$

$$\lambda_4 = 0.04036$$

(v)

Hazard function for a cigarette smoker an any duration 'x'

$$\lambda_x = (e^{\beta_0+\beta_1 x_1+\beta_2}) * c^x$$

Hazard function for a non-smoker at any time 't'

$$\lambda'_t = (e^{\beta_0+\beta_1 x_1}) * c^t$$

Assuming both the hazard functions are equal,

$$\lambda_x = \lambda'_t$$

$$(e^{\beta_0+\beta_1 x_1+\beta_2}) * c^x = (e^{\beta_0+\beta_1 x_1}) * c^t$$

$$(e^{\beta_2}) * c^x = c^t$$

$$(e^{0.65}) = c^{t-s}$$

$$1.915541 = c^{t-s}$$

$$\ln(1.915541) = (t - s)\ln(1.05)$$

$$t - s = 13.32$$

This proves that both the hazard functions will be constant when the difference between the age of a non-smoker and that of smoker is approximately 13.32 years.

13.

(i)

For age 56

$$u_x = \frac{d_x}{E_x^c}$$

Let dx be the number of deaths at age x nearest birthday

Let $P_{x,t}$ = number of policies in-force for age x nearest birthday at duration t

Let $P'_{y,t}$ = number of policies in-force for age y last birthday

$$E_{56}^c = \int_0^2 P_{56,t} dt$$

$$\int_0^2 P_{56,t} dt = \frac{1}{2}\left(P_{56,0} + P_{56,1}\right) + \frac{1}{2}\left(P_{56,1} + P_{56,2}\right)$$

$$P_{56,0} = \frac{1}{2} * \left(P'_{55,0} + P'_{56,0}\right) = \frac{1}{2} * (20100 + 20000) = 20050$$

$$P_{56,1} = \frac{1}{2} * \left(P'_{55,1} + P'_{56,1}\right) = \frac{1}{2} * (20500 + 21100) = 20800$$

$$P_{56,2} = \frac{1}{2} * \left(P'_{55,2} + P'_{56,2}\right) = \frac{1}{2} * (18500 + 20000) = 19250$$

$$\int_0^2 P_{56,t} dt = \frac{1}{2}\left(P_{56,0} + P_{56,1}\right) + \frac{1}{2}\left(P_{56,1} + P_{56,2}\right)$$

$$\int_0^2 P_{56,t} dt = \frac{1}{2} * P_{56,0} + P_{56,1} + \frac{1}{2} * P_{56,2}$$

$$E_{56}^c = 40450$$

$$u_{56} = \frac{d_{56}}{E_{56}^c} = \frac{1380}{40450} = 0.0341162$$

For age 57

$$u_x = \frac{d_x}{E_x^c}$$

Let dx be the number of deaths at age x nearest birthday

Let $P_{x,t}$ = number of policies in-force for age x nearest birthday at duration t

Let $P'_{y,t}$ = number of policies in-force for age y last birthday

$$E_{57}^{c} = \int_{0}^{2} P_{57,t} dt$$

$$\int_{0}^{2} P_{57,t} dt = \frac{1}{2}\left(P_{57,0} + P_{57,1}\right) + \frac{1}{2}\left(P_{57,1} + P_{57,2}\right)$$

$$P_{57,0} = \frac{1}{2} * \left(P'_{56,0} + P'_{57,0}\right) = \frac{1}{2} * (20000 + 19700) = 19850$$

$$P_{57,1} = \frac{1}{2} * \left(P'_{56,1} + P'_{57,1}\right) = \frac{1}{2} * (21100 + 20700) = 20900$$

$$P_{57,2} = \frac{1}{2} * \left(P'_{56,2} + P'_{57,2}\right) = \frac{1}{2} * (20000 + 15000) = 17500$$

$$\int_{0}^{2} P_{57,t} dt = \frac{1}{2}\left(P_{57,0} + P_{57,1}\right) + \frac{1}{2}\left(P_{57,1} + P_{57,2}\right)$$

$$\int_{0}^{2} P_{57,t} dt = \frac{1}{2} * P_{57,0} + P_{57,1} + \frac{1}{2} * P_{57,2}$$

$$E_{57}^{c} = 39575$$

$$u_{57} = \frac{d_{57}}{E_{57}^{c}} = \frac{1420}{39575} = 0.03588124$$

(ii)

$$q_{x-\frac{1}{2}} = (1 - e^{-u_x})$$

$$q_{55.5} = (1 - e^{-u_{56}}) = (1 - e^{-0.0341162}) = 0.033541$$

And,

$$q_{56.5} = (1 - e^{-u_{57}}) = (1 - e^{-0.03588124}) = 0.03524514$$

14.

(i)

Expression for the hazard function for the model specified

$$h_i(t) = h_0(t) * e^{(0.3*P1+0.5*P2-0.1*P3+0.3*G+0.2*L+0.7*A1+0.5*A2-0.4*A3)}$$

Where,

$h_i(t)\ is\ the\ hazard\ function\ of\ the\ i^{th} life\ after\ searching\ for\ time\ t\ for\ a\ life\ partner$

$h_0(t)\ is\ the\ baseline\ hazard\ function\ of\ the\ i^{th} life\ which\ depends\ only\ on\ t$

$P1 = 1\ if\ service\ else\ 0\ if\ other,$

$P2 = 1, if\ business\ else\ 0\ if\ other,$

$P3 = 1\ if\ social\ service\ else\ 0\ if\ other,$

$G = 1\ if\ female\ else\ 0\ if\ male,$

$L = 1\ if\ non\ metro\ else\ 0\ if\ metro,$

$A1 = 1\ if\ 20 - 25\ else\ 0\ if\ other,$

$A2 = 1\ if\ 25 - 30\ else\ 0\ if\ other$

$A3 = 1\ if\ 35 - 40\ else\ 0, if\ other$

$e^{(0.3*P1++0.5*P2-0.1*P3+0.3*G+0.2*L+0.7*A1+0.5*A2-0.14*A3)}$

$is\ the\ effect\ of\ covariates\ to\ the\ hazrad\ function$

(ii)

The probability for staying single for next two years for a female social worker aged 37 living in Mumbai who has been looking to get married since last 1 year is 0.3

$$\therefore 0.3 = e^{-\int_1^3 h_{social\,service, female, metro, 35-40}(t)dt}$$

$$0.3 = e^{-\int_1^3 h_0(t)*e^{(-0.1+0.3-0.4)}dt}$$

$$0.3 = e^{-0.818731\int_1^3 h_0(t)dt}$$

Taking ln

$$\ln(0.3) = -0.818731\int_1^3 h_0(t)dt$$

$$\int_1^3 h_0(t)dt = 1.4705353$$

To find: The probability for staying single for next two years for a male aged 24 living in non-metropolitan city and doing business given that he is looking for a partner for 1 year

$$\therefore Required\ probability = e^{-\int_1^3 h_{business,male,non-metro,20-25}(t)dt}$$

$$= e^{-\int_1^3 h_0(t)*e^{(0.5+0.2+0.7)}dt}$$

$$= e^{-4.0552*1.4705353}$$

$$\therefore Required\ probability = 0.0025714$$

The probability for staying single for next two years for a male aged 24 living in non-metropolitan city and doing business given that he is looking for a partner for 1 year is 0.0025714.

15.

(i)

Advantages of using central exposed to risk in actuarial investigations as opposed to initial exposed to risk are as follows:

1] Central exposed to risk makes more practical sense as opposed to initial exposed to risk since, central exposed to risk takes into account the duration which is actually spent by the observation at risk.

2] The central force of mortality (u_x) is easier to interpret in the case of central exposed to risk rather than initial exposed to risk.

(ii)

Rita will attain exact age 30 last birthday on 1st October 2009.

Central exposed to risk of Rita at age 30 last birthday (E_{30}^{c}) = 1st October 2009 to 31st December 2009 i.e., 3 months

Initial exposed to risk of Rita at age 30 last birthday = 1st October 2009 to 30th September 2010 i.e., 1 year

Sita will attain exact age 30 last birthday on 1st September 2011.

Central exposed to risk of Sita at age 30 last birthday (E_{30}^{c}) = 1st September 2011 to 31st August 2012 i.e., 1 year

Initial exposed to risk of Sita at age 30 last birthday = 1st September 2011 to 31st August 2012 i.e., 1 year

Nita will attain exact age 30 last birthday on 1st December 2009.

Central exposed to risk of Nita age 30 last birthday (E_{30}^{c}) = 1st February 2010 to 31st October 2010 i.e., 9 months

Since, she was no more married, Nita is censored

Thus, initial exposed to risk of Nita at age 30 last birthday is the same i.e., 9 months.

Gita will attain exact age 30 last birthday on 1st April 2010.

Since, she was married at age 31 last birthday, therefore, her contribution to central and initial exposed to risk is 0

(iii)

Total central exposed to risk

$$= \frac{3}{12} + 1 + \frac{9}{12} + 0 = 2\ years$$

Total initial exposed to risk

$$= 1 + 1 + \frac{9}{12} = 2.75\ years$$