MTH6157 Assessed Coursework 2

Please paste your R code and output for parts (a), (b) and (c) into a Word document along with your typed answers to parts (d) and (e) and submit that Word document via the submission point in QM Plus by 5pm UK time on Friday 24 November 2023 (week 9). Please make sure that you submit your own work and that you are aware of QMUL policies on collusion and plagiarism.

<u>Question</u>

Three small life assurance companies: Amber Life, Diamond Inc and Sapphire plc agree to cooperate and combine their recent mortality experience for the modelling of mortality rates. All three companies use a 2-State model. The data for this analysis is found in three CSV files on QM Plus, Amber.csv, Diamond.csv and Sapphire.csv each of which contains deaths and number of insurance policies by age. The IT systems of these three companies are all different. The data in each CSV file is summarised below:

Amber
4 columns with the number of deaths in 2022 and number of policies on 1 January 2022, 30 June 2022 and 31 December 2022. Age defined as age last birthday.

Diamond
14 columns with the number of deaths in 2022 and number of policies on 1 January 2022 and then at the end of each month in 2022. Age defined as age last birthday.

Sapphire
3 columns with the number of deaths in 2022 and number of policies on 1 January 2022 and 31 December 2022. Age defined as age next birthday.

(a) Import these three csv files into R assigning each column of data to a suitably named vector.[2]

```
> Amber <- read.csv("~/Amber.csv")
> agelast <- Amber$age
> dA <- Amber$deaths
> PA1 <- Amber$January1
> PA2 <- Amber$June30
> PA3 <- Amber$December31
> # assign d to deaths and P to policy census
> #dA and PA for Amber
> #PA1 PA2 PA3 for 3 census dates from Amber
> Diamond <- read.csv("~/Diamond.csv")
> dD <- Diamond$deaths
> PD1 <- Diamond$Jan1
> PD2 <- Diamond$Jan31
> PD3 <- Diamond$Feb28
> PD4 <- Diamond$Mar31
> PD5 <- Diamond$Apr30
> PD6 <- Diamond$May31
> PD7 <- Diamond$Jun30
> PD8 <- Diamond$Jul31
> PD9 <- Diamond$Aug31
> PD10 <- Diamond$Sep30
> PD11 <- Diamond$Oct31
> PD12 <- Diamond$Nov30
> PD13 <- Diamond$Dec31
> # use dD and PD for Diamond deaths and census
> # monthly data so PD1 to PD13
> Sapphire <- read.csv("~/Sapphire.csv")
> agenext = Sapphire$age
```

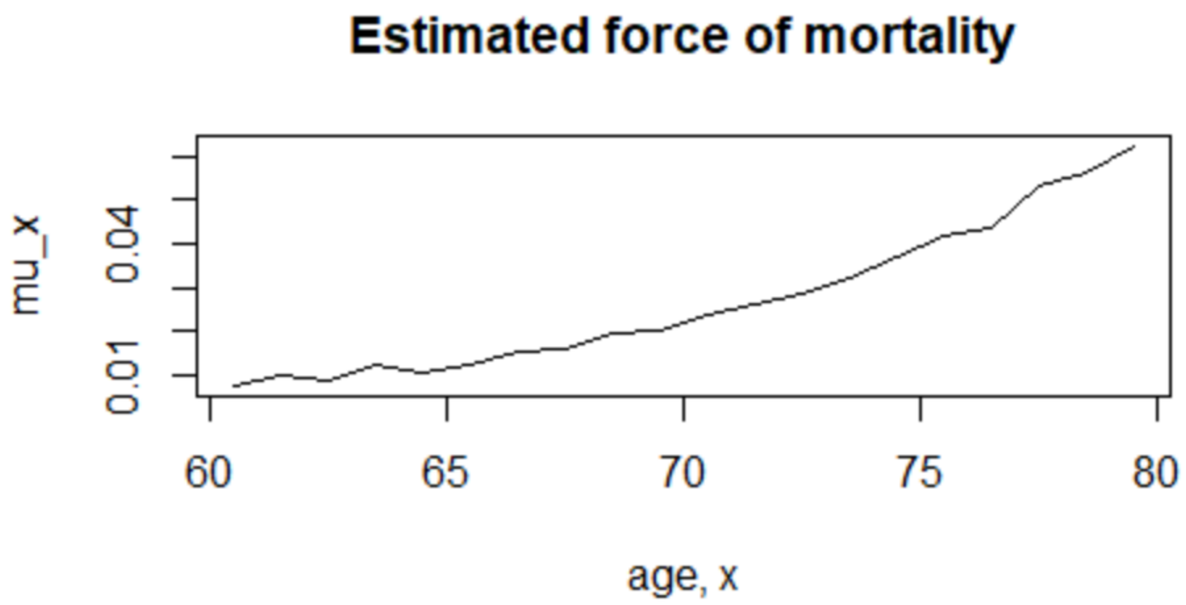
```
> dS <- Sapphire$deaths
> PS1 <- Sapphire$January1
> PS2 <- Sapphire$December31
> #dS and PS1 PS2 for Sapphire
> # all the data now imported and assigned
```

(b) Calculate and display estimates of the force of mortality at ages 60½, 61½, 62½, … 79½ combining the data from the three insurance companies. Include details of any assumptions you make as comments in your R script. [15]

```
> # we can sum deaths across the three companies
> # as 2 of 3 use age last we will adopt that
> deaths = dA + dD + dS
> # we need to calculate exposed to risk separately
> # for each company because of different census dates
> # we use EA ED ES for each company's E-to-R
> EA = 0.5 * (0.5*PA1 + PA2 + 0.5*PA3)
> ED = (1/12) * (0.5*PD1 + PD2 + PD3 + PD4 + PD5 + PD6 + PD7 + PD8 + PD9
+ PD10 + PD11 + PD12 + 0.5*PD13)
> ES = 0.5 * (PS1 + PS2)
> # then sum the three company E-to-R for
> # total exposed to risk
> exposed = EA + ED + ES
> # the age next values of Sapphire align with the age last values of the
other two hence no adjustment actually needed
> # finally mu_hat at required age is deaths/ exposed
> mu_hat = deaths / exposed
> # now to display the results use cbind() to combine agelast and mu_hat
in a data frame
> # the mu-hat age is age last + 1/2
> age = agelast + 0.5
> table = cbind(age, mu_hat)
> table
       age      mu_hat
 [1,] 60.5 0.007352224
 [2,] 61.5 0.009530022
 [3,] 62.5 0.008567545
 [4,] 63.5 0.011947729
 [5,] 64.5 0.010337497
 [6,] 65.5 0.011867969
 [7,] 66.5 0.015181309
 [8,] 67.5 0.016079325
 [9,] 68.5 0.019706420
[10,] 69.5 0.019960830
[11,] 70.5 0.023554781
[12,] 71.5 0.025888177
[13,] 72.5 0.028738604
[14,] 73.5 0.032171582
[15,] 74.5 0.037007424
[16,] 75.5 0.042088791
[17,] 76.5 0.043844110
[18,] 77.5 0.053433045
[19,] 78.5 0.056370340
[20,] 79.5 0.062450162
```

(c) Plot the estimates calculated in (b) above by age in a suitably labelled graph. [4]

```
> plot(age, mu_hat, type = "l", main = "Estimated force of mortality",
xlab = "age, x", ylab = "mu_x")
```



(d) Comment briefly on the main features of the graph in (c) above. [3]

- The force of mortality appears to increase exponentially with age
- Which is what we would expect
- However the mu_x do not proceed smoothly with age
- So more work is needed before using these in practical work
- E.g. mu_62½ > mu_63½ which is not what we expect
- The rate of increase varies particularly at the oldest ages

(e) What are the advantages and disadvantages of combining the data of these three companies into a single set of estimates. [6]

- The main advantage is the pooling of data increases deaths and exposure amounts
- Which should lead to lower variance estimates of mu_x
- Especially at older ages where we expect fewer policies in general
- However the disadvantages relate to how selection might influence the results
- Class selection and temporary initial selection are the 2 important types here
- Class selection might be present in different ways across the three companies
- Most importantly in different mixes of protection and annuity business
- But potentially in other ways that impact homogeneity
- Temporary initial selection effects might differ between the companies
- Both in the amount of selection and the select period
- If underwriting practices differ between the three companies
- Spurious selection may also be present
- More investigation of the underlying policy data is needed to understand how this might affect modelling results

[Total 30]

Your mark out of 30 will be converted into a percentage to be compatible with the mark from the first assessed coursework.