Queen Mary
University of London

Main Examination period 2022 – May/June – Semester B

MTH5131: Actuarial Statistics

You should attempt ALL questions. Marks available are shown next to the questions.

In completing this assessment:

- **You may use books and notes.**
- **You may use calculators and computers, but you must show your working for any calculations you do.**
- **You may use the Internet as a resource, but not to ask for the solution to an exam question or to copy any solution you find.**
- **You must not seek or obtain help from anyone else.**

All work should be **handwritten** and should **include your student number**.

The exam is available for a period of **24 hours**. Upon accessing the exam, you will have **3 hours** in which to complete and submit this assessment.

When you have finished:

- scan your work, convert it to a **single PDF file**, and submit this file using the tool below the link to the exam;
- e-mail a copy to **maths@qmul.ac.uk** with your student number and the module code in the subject line;
- with your e-mail, include a photograph of the first page of your work together with either yourself or your student ID card.

Please try to upload your work well before the end of the submission window, in case you experience computer problems. **Only one attempt is allowed – once you have submitted your work, it is final**.

IFoA exemptions. For actuarial students, this module counts towards IFoA actuarial exemptions. To be eligible for IFoA exemption, **you must submit your exam within the first 3 hours of the 24-hour exam period**.

Examiners: D. S. Stark, S. D. Coad

Question 1 [10 marks]. The registrar of Queen Mary wants to estimate how many students are living at home. Determine whether the following schemes of sampling students are simple random sampling, stratified sampling or neither. Explain your answers.

(a) The registrar keeps an alphabetical list of students. It is proposed that a number be chosen at random from 1 to 100 and the name that far down in the list and every hundredth name after it be selected. **[4]**

(b) It is proposed that each school in the university chooses a predetermined number of randomly chosen students. **[3]**

(c) It is proposed that someone should stand in front of the library and select a random sample of entering students. **[3]**

Question 2 [12 marks]. Consider a 6×2 data matrix

$$\begin{pmatrix} 19 & 12 \\ 22 & 6 \\ 6 & 9 \\ 3 & 15 \\ 2 & 13 \\ 20 & 5 \end{pmatrix}$$

of 6 observations of 2 variables.

(a) Find the sample covariance matrix of the data. **[5]**

(b) Find the component of the data corresponding to the largest variance. **[7]**

Question 3 [16 marks]. The following data shows a company's production and its employees' aggregate salaries in millions of pounds over 5 years.

Production	1000	2000	2500	4000	2300
Salaries	190	200	250	700	180

(a) (i) Compute the Pearson correlation coefficient r_s between production and salaries, showing your calculations and without using the R cor function. **[5]**

(ii) Under the assumption that the data are bivariate normally distributed with correlation parameter ρ, test $H_0 : \rho = 0$ against $H_1 : \rho \neq 0$ at the 5% level, showing your calculations and without using the R cor.test function. **[5]**

(b) Compute the Kendall correlation coefficient between production and salaries, showing your calculations and without using the R cor function. **[6]**

Question 4 [13 marks]. Suppose that $Y_1, Y_2, \ldots, Y_n$ are independent $N(5, \theta)$ random variables, where $\theta > 0$.

(a) Find the Cramér-Rao lower bound for unbiased estimators of θ. **[7]**

(b) Given that $\text{Var}((Y_i - 5)^2) = 2\theta^2$, show that

$$\frac{(Y_1 - 5)^2 + (Y_2 - 5)^2 + \cdots + (Y_n - 5)^2}{n}$$

is an unbiased estimator of θ of smallest variance. **[6]**

Question 5 [13 marks]. Let $Y_1, Y_2, \ldots, Y_n$ be independent and identically distributed with probability density function

$$f_Y(y) = \begin{cases} \theta y^{\theta-1} & \text{for } 0 < y < 1, \\ 0 & \text{otherwise,} \end{cases}$$

where $\theta > 0$ is a parameter.

(a) Show that $\overline{Y} := \frac{Y_1+Y_2+\cdots+Y_n}{n}$ is a consistent estimator for $\frac{\theta}{\theta+1}$. **[6]**

(b) Determine whether $\frac{n}{n+1}\overline{Y}$ is a consistent estimator for $\frac{\theta}{\theta+1}$. **[7]**

Question 6 [11 marks]. The Department of Agriculture and Rural Development (DARP) is currently seeking foxes in order to test them for Trichinella, a disease found in pigs. Suppose that out of 24 foxes, 3 of them are found to have Trichinella. Let θ be the probability that a fox has Trichinella. Our prior distribution of θ is Beta(1.5, 1.5).

(a) What is the posterior distribution of θ? **[7]**

(b) Find the Bayesian estimate of θ under all-or-nothing loss. **[4]**

Question 7 [13 marks]. Suppose the pure premium arising from a risk each year, measured in hundreds of thousands of pounds, has a $\mathrm{N}(\theta, 400)$ distribution, where the prior distribution for θ is $\mathrm{N}(700, 500)$. You have been given data as shown in the first two columns in the table below.

Year	Pure Premium	Credibility factor at the start of year	Average pure premium based on number of years of past data available at the start of year	At the start of the year, the credibility estimate of the pure premium in the coming year
1	760			
2	735			
3	790			

Using the data from the first two columns, calculate the remaining columns. **[13]**

Question 8 [12 marks]. A statistician obtains data for eight houses. Variables for each house are selling price (in thousands of pounds), size of house (in square feet), annual property tax bill (in pounds), number of bedrooms, number of bathrooms, and the factor New which is 0 if the house is old and 1 if it is new.
The statistician runs the following R code and obtains the given output.

```
---------------------------------------------------------------------
> fit.gamma <- glm(price ~ size + new + beds + size:new + size:beds,
                   family = Gamma(link = identity))
> summary(fit.gamma)$coef
              Estimate  Std. Error  t value   Pr(>|t|)
(Intercept)    44.3759    48.5978    0.9131    0.3635
size            0.0740     0.0400    1.8495    0.0675
new            60.0290    65.7655   -0.9128    0.3637
beds           22.7131    17.6312   -1.2882    0.2008
size:new        0.0538     0.0376    1.4325    0.1553
size:beds       0.0100     0.0126    0.7962    0.4279
---------------------------------------------------------------------
```

(a) What model is being fitted? Give a complete answer stating all three components of the model. **[3]**

(b) If you were selecting an optimal model by backward selection, which model would you check next? Justify your answer. **[4]**

(c) Suppose that a house has 2064 square feet, three bedrooms, and is old. What does the fitted model predict that the selling price is? **[5]**

End of Paper.