Queen Mary
University of London

Main Examination period 2021 – May/June – Semester B
Online Alternative Assessments

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**.

You have **24 hours** 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.

You are expected to spend about 2 hours to complete the assessment, plus the time taken to scan and upload your work. 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. You are allowed two submissions for this exam—the first for your IFoA mark, and the second for your module mark. To be eligible for IFoA exemptions, **your IFoA submission must be within the first 3 hours of the assessment period**.

Examiners: Dr. D. Stark, Dr. C. Sutton

Question 1 [12 marks]. You are assigned the task of studying whether COVID-19 mitigation measures, including the lockdown, may have unintended health consequences. It was decided that the use of alcohol, cigarettes and e-cigarettes would be compared before and after the lockdown.

(a) Explain what sources of data you might use for this study. **[3]**

(b) Explain what form of data analysis this study is (i.e. descriptive, inferential or predictive). **[3]**

(c) Explain what cross-sectional and what longitudinal data is relevant to this study. **[3]**

(d) Explain what data visualisations might be used in this study. **[3]**

Question 2 [9 marks]. Suppose three tests are administered to a random sample of college students. Let $\hat{X}$ be a $n \times 3$ matrix, where the score of the ith students on the jth test is $(\hat{X})_{i,j}$. Suppose that the sample covariance matrix of $\hat{X}$ is

$$S = \begin{pmatrix} 5 & 2 & 0 \\ 2 & 6 & 2 \\ 0 & 2 & 7 \end{pmatrix}.$$

The eigenvalues of S are 3, 6, and 9.

(a) Find the component corresponding to the eigenvalue 9. **[6]**

(b) There are components corresponding to each of the eigenvalues. Which of them would you say are principal? Explain your reasoning. **[3]**

Question 3 [16 marks]. Two interviewers ranked 15 candidates (A through O) for a position. The results from most preferred to least preferred are:

Interviewer1 : ABCDEFGHIJKLMNO

Interviewer2 : CABDFEHGJILKONM

The rankings of the candidates by the two interviewers were therefore

Candidate	Interviewer 1	Interviewer 2
A	7	8
B	2	3
C	10	9
D	14	14
E	15	13
F	6	5
G	1	2
H	8	7
I	9	10
J	4	4
K	5	6
L	3	1
M	13	15
N	11	12
O	12	11

(a) (i) Compute the Spearman correlation coefficient r_s between the two rankings. **[5]**

(ii) For a t-distribution approximation involving r_s, find the p-value of the t-statistic and use it to decide whether or not the rankings are correlated. **[4]**

(b) Compute the Kendall correlation coefficient between the two rankings. **[7]**

Question 4 [18 marks]. Let Y denote the proportion of allotted time that a randomly selected student spends working on a certain aptitude test. Suppose that Y is $\text{Beta}(\theta, 1)$ distributed, where $\theta > 0$. A random sample of ten students yields the data

$$0.92,\ 0.79,\ 0.90,\ 0.65,\ 0.86,\ 0.47,\ 0.73,\ 0.97,\ 0.94,\ 0.77$$

(a) Obtain the method of moments estimate of θ for this data. **[6]**

(b) Obtain the maximum likelihood estimate of θ for this data. **[5]**

Suppose that in addition to the random sample of ten students, you know that two other random students took less than 60% of their allotted time working on the aptitude test.

(c) Obtain the maximum likelihood estimate of θ for this data and comment on the difference between the maximum likelihood estimate of θ that you obtain and the maximum likelihood estimate of θ that you obtained in part (b). **[7]**

Question 5 [12 marks]. Let $Y_1 \sim \text{Exponential}(\alpha)$, $Y_2 \sim \text{Exponential}(2\alpha)$ and $Y_3 \sim \text{Exponential}(4\alpha)$ be independent, where $\alpha > 0$ is a parameter.

(a) Determine a constant C such that $C(Y_1 + Y_2 + Y_3)$ is an unbiased estimator for $1/\alpha$. **[7]**

(b) Find the Mean Square Error of the unbiased estimator you found in part (a). **[5]**

Question 6 [9 marks]. Suppose that $Y_1, Y_2, \ldots, Y_{10}$ are i.i.d. $\text{Geometric}(p)$ with common p.m.f.

$$P(Y = y) = p(1-p)^y, \quad y = 0, 1, 2, \ldots$$

The prior distribution on p is $\text{Beta}(4, 3)$.

(a) Given that the data $y_1, y_2, \ldots, y_n$ satisfy $\sum_{i=1}^{10} y_i = 42$, find the posterior distribution. **[6]**

(b) Find the Bayesian estimate of p under squared error loss. **[3]**

Question 7 [9 marks]. The table below shows aggregate annual claim statistics for four different products over a period of five years. Annual aggregate claims for product i in year j are denoted by X_{ij}.

Product (i)	$\overline{X_i} = \frac{1}{5}\sum_{j=1}^{5} X_{ij}$	$\frac{1}{4}\sum_{j=1}^{5}(X_{ij} - \overline{X_i})^2$
1	125	130
2	85	60
3	140	35
4	175	100

(a) Calculate the credibility premium of each product under the assumptions of EBCT Model 1. **[6]**

(b) Find the credibility estimate of the amount per claim for the coming year for product 3. **[3]**

Question 8 [15 marks]. A small insurer wishes to model its number of claims for motor insurance using a simple generalised linear model based on the two factors:

$$YO_i = \begin{cases} i = 1 & \text{for 'young' drivers;} \\ i = 0 & \text{for 'old' drivers;} \end{cases}$$

$$FMS_j = \begin{cases} j = 2 & \text{for 'fast' cars;} \\ j = 1 & \text{for 'medium' cars;} \\ j = 0 & \text{for 'slow' cars;} \end{cases}$$

The data is the 6 aggregate claim costs for each combination of factors.
The insurer is considering three possible models for the linear predictor, a constant model plus the two models specified below:

$$\text{Model 1}: \quad YO$$

$$\text{Model 2}: \quad YO + FMS$$

The Poisson distribution is used with its canonical link.

(a) Write each of these models in parameterised form, stating how many non-zero parameter values are present in each model. **[6]**

(b) The table below shows the students calculated values of the scaled deviance for these two models and the constant model. The degrees of freedom is defined to be the amount of data minus the number of non-zero parameters fitted.

Model	Scaled Deviance	Degrees of Freedom
1	40	5
YO	30	
$YO + FMS$	23	

Complete the table by filling in the missing entries in the degrees of freedom column and carry out the χ^2 tests and calculations of change in AIC to determine which model would be the most appropriate. When applying a χ^2 test, use significance level 5%. **[9]**

End of Paper.