MA2261
All candidates

June Examinations 2022

DO NOT OPEN THE QUESTION PAPER UNTIL INSTRUCTED TO DO SO BY THE CHIEF INVIGILATOR	
School	School of Computing and Mathematical Science
Module Code	MA2261
Module Title	Linear Statistical Models
Exam Duration	2 hours
CHECK YOU HAVE THE CORRECT QUESTION PAPER	
Number of Pages	9
Number of Questions	4
Instructions to Candidates	Exam is marked out of 60 points. Answer all questions.
FOR THIS EXAM YOU ARE ALLOWED TO USE THE FOLLOWING:	
Calculators	Approved calculators may be used.
Books/Statutes provided by the University	Statistical tables are provided
Are students permitted to bring their own Books/Statutes/Notes?	No
Additional Stationery	No

1. A continuous random variable X has a probability density function given by

$$f(x,\lambda) = \begin{cases} \dfrac{xe^{-(\frac{x}{\lambda})}}{\lambda^2}, & x > 0 \\ 0, & x \leq 0\,, \end{cases}$$

where λ is a parameter and $\lambda > 0$.

In a factory of electronic components, the life time X is assumed to follow the above distribution. Twenty components were randomly selected, tested and their following life time observations $x_1, \ldots x_{20}$ of X (in thousand hours) were recorded.

16	6	1.2	4.2	3.2	7.5	11	16	7.5	21
6.5	2	18	14	9.5	13.2	13.9	6.2	12.2	31

In your answers you can use, without proof, the following calculus facts:

$$\frac{d}{dz}\left(-e^{-z}(z+1)\right) = ze^{-z},$$

$$\sum_{i=1}^{20} x_i = 220.1$$

a) **[5 marks]**

Find the cumulative distribution function of X.

b) **[5 marks]**

Show that the log-likelihood function $\ell(\lambda)$ for the given dataset is

$$\ell(\lambda) = \ln \mathcal{L}(\lambda) = -2n\ln\lambda - \sum_{i=1}^{n}\frac{x_i}{\lambda} + \sum_{i=1}^{n}\ln x_i$$

c) **[5 marks]**

Apply the method of maximum likelihood to find the maximum likelihood estimate of the parameter λ for the given dataset. You may assume any critical point of the log-likelihood function is a maximum.

Total 15 marks

2. A team of scientists from the University of Oxford is on a mission in Antarctica to study the dust concentration (Y) in the ice of different geological eras. To this end, they drill the ice and obtain samples at different depths (between 300 and 1600 m) corresponding to periods (x) in the past between 3000 and 280,000 years ago. The collected data and statistical analysis, using simple linear regression, were sent back to Oxford. kYr means 1000 years.

$\bar{x} = 92.4$, $\bar{y} = 86.805$, $S_{xx} = 85,114.4$, $S_{yy} = 19,179.1$, $S_{xy} = 33,123.77$, $RSS = 6,288.401$

x (kYr)	Y Dust concentration (ppm)
3	39.36
6	24.25
15	36.91
25	97.32
46	60.5
66	114.91
102	81.5
161	121.3
220	164.8
280	127.2

The Oxford team has a deadline to submit their findings to a Government agency and they need to produce, without delay, a statistical analysis using the information they have.

a) **[3 marks]**
Obtain the estimated simple linear regression line.

b) **[10 marks]**
Decide by conducting a statistical test with significance level $\alpha = 5\%$ if the period variable x has a statistically significant effect on the mean dust concentration.

c) **[2 marks]** Calculate a point estimate for the mean increase in dust concentration corresponding to an increase in one unit (that is 1000 years) in x.

Total: 15 marks

3. Hooke's law states that the elongation L of a spring subjected to a weight force W is given by $W = kL$, where W is measured in kg, L is measured in mm, k is called the elastic constant of the spring and is expressed in kg/mm. A group of Physics students wants to measure the elastic constant of a spring. Not having very precise instruments they perform an experiment by applying to the spring weights of increasing intensity, from 10 to $50\,kg$, for five times. The measurements are influenced by the approximation of the reading of the lengths and by the fact that the spring does not behave like a perfect spring and the same weight applied several times does not give the same elongation. Therefore, we assume that L is normally distributed for a given weight W. The measurements of elongation, in millimeters, for each test done are shown below.

Weight (kg)	Measured elongation L (mm)				
W	L1	L2	L3	L4	L5
10	48.6	47.6	48.8	51.5	49.8
15	78.4	77.5	71.6	77.5	73.6
20	95.7	98.6	100.4	102.4	97.3
25	123.5	131.1	118.9	130.6	128.3
30	150.6	154.5	148.3	146.0	153.3
35	175.5	176.3	173.2	181.8	181.8
40	209.4	199.8	197.8	195.9	203.5
45	230.9	233.2	230.5	218.7	222.6
50	245.2	249.2	257.0	256.7	244.3

a) **[4 marks]**

Write the equation of a simple linear regression model with repeated observations for this dataset.

b) **[8 marks]**

Perform a statistical test with significance level $\alpha = 5\%$ to establish if a linear regression model with repeated observations is a good fit for these data. You can use the following values:

$SST = 192093$, $SSM = 191345$ and $SSB = 191387$.

c) **[3 marks]**

Write the general formula of the test statistic and its distribution you can use to decide if Hooke's law can be assumed to be valid for this dataset.

Total: 15 marks

4. In this question, X and Y are jointly distributed discrete random variables. The probability $P(X=i, Y=j)$ is given by the entry in column i, row j of the following table with $p, q \in [0,1]$:

		X = 0	X = 1	X = 2
Y	0	0	pq	0
	1	$\frac{p(1-q)}{2}$	$1-p$	$\frac{p(1-q)}{2}$

a) **[5 marks]**
Calculate $P(Y=0)$ and $P(Y=1)$.

b) **[8 marks]**
For which values of p and q are the **events** $\{X=0\}$ and $\{Y=1\}$ independent?

c) **[2 marks]**
Give a formal definition of what it means for two random variables to be independent.

Total: 15 marks