

CM1606: Computational Mathematics Assignment – Statistics (2021 Jan Intake)	
Module Leader	Ganesha Thondilage
Stage	1
Unit (Group/Individual)	individual
Weighing	60%
Qualifying Mark	D grade
Learning Outcomes Covered in this Assignment:	Data representation, analysis, and visualization; application of statistical models and hypothesis testing to real life problems
Handed Out Date	28 May 2021
Due Date	09 Jul 2021
Expected Deliverable	Softcopy
Method of Submission	Online through Campus Moodle
Method of Feedback and Due Date	Written/rubric based; Jul 21
BCS Criteria(Pending) Met by this Assignment	

### Assessment Regulations

Refer to the “How you are assessed section” in the Student Handbook for undergraduate students for a clarification of how you are assessed, penalties and late submissions, what constitutes plagiarism etc.

### Penalty for Late Submission

Coursework received late without valid reason shall not be accepted and shall receive no grade, but shall count as one of the assessment opportunities prescribed in paragraph 9 of **RGU Academic Regulation A4 section 4.3**.

It is recognized that on occasion, illness, personal crisis or other valid circumstances can mean that you fail to submit and/or attend an assessment on time. In such cases you must inform the School of any extenuating circumstances through a **Coursework Extension Form** or a **Deferral Request Form**, with valid evidence for non-submission of an assessment up to a maximum of five working days after the assessment submission date. This information will be reported to the relevant Assessment Board that will decide whether a student should be allowed to reattempt without penalty (a deferral). For more detailed information regarding University Assessment Regulations and accessing forms, please refer to the following website: [www.rgu.ac.uk/academicregulations](http://www.rgu.ac.uk/academicregulations)

### Grading

Marks will be awarded for the coursework based on the provided Grading Grid. These marks will be mapped onto a grade scale from A-F as determined by the individual module coordinator.

Instructions to Candidates:

1. There are 6 questions in the assignment. Answer all questions.
2. Each question is worth 10 marks with a breakdown as shown.
3. You may use R software for calculations and plots.

**QUESTION 1.**

A class of 25 students got the following marks for mathematics:

36, 39, 40, 37, 40, 50, 51, 51, 43, 45, 53, 56, 61, 62,  
46, 48, 41, 44, 64, 74, 82, 72, 75, 90, 66

1. Make a histogram of the data using the seven bins 30-39; 40-49; 50-59; 60-69; 70-79; 80-89; 90-99. [don't have gaps between bars].  
(2.5 marks)

2. Comment on the shape of the distribution and what it means.  
(2.5 marks)

3. Calculate the mean and find the median.  
(2.5 marks)

4. What do you think is a more suitable measure of center, mean or median, for this class? Justify your answer.

(2.5 marks)

(Total 10 marks)

## QUESTION 2.

An interview was held to select three software engineers for a company. Five applicants (A, B, C, D, E) were shortlisted after a preliminary interview. Panel of the 2<sup>nd</sup> interview consisted of three members: X: Company Managing Director (MD), Y: Information Systems Manager, and Z: Computer Scientist from outside.

The MD evaluated applicants based on overall performance whereas other two evaluated mostly based on subject knowledge. The following are the ratings given by the panel: (1 = best; 5 = worst)

<u>Applicant</u>	<u>Panel</u>		
	X	Y	Z
A	3	1	1
B	1	3	2
C	4	2	3
D	2	5	4
E	5	4	5

1. Calculate the degree of pairwise agreement between panel members using an appropriate measure. (4 marks)

2. Comment on the results of 1. (3 marks)

3. What's your opinion about the choice of the best three applicants? (this must be based on the ratings given)

(3 marks)

(Total 10 marks)

**QUESTION 3.**

Consider an Ethernet connection with capacity of 100 Mbit per second.

Let  $Y$  = Number of Mbit of data transmitted per second.

Assume  $Y \sim \text{Poisson } (\lambda=80)$ .

Calculate the probability of utilizing:

1. 85% of the capacity of the channel. **(5 marks)**

2. 90% of the capacity of the channel. **(5 marks)**

**(Total 10 marks)**

**QUESTION 4.**

Customers arrive at a service counter according to a Poisson distribution with rate  $\lambda = 20$  per hour. Calculate the probability that:

1. Service completion time for any customer is less than 3 minutes. **(5 marks)**

2. Service completion time for any customer is less than 4 minutes. **(5 marks)**

**(Total 10 marks)**

**QUESTION 5.**

Consider a sample of three values:  $\{2, 3, 4\}$ .

1. Write down all possible distinct bootstrap samples **(4 marks)**

2. Find the median of each bootstrap sample. **(3 marks)**

3. Estimate (by hand) the 'bootstrap bias' of the sampling distribution of the sample median. (use only the samples in 1).

**(3 marks)**

**(Total 10 marks)**

### QUESTION 6.

The following data shows the annual license fee (Rs. Mn) for 2019 and 2020 based on total assets at the end of 2018, for some licensed commercial and specialized banks in Sri Lanka [Source: Annual Report, Central Bank, 2018].

Total assets (Rs. Bn), x	75	125	200	500	750	1000
License fee (Rs. Mn), y	6	12	17	23	26	32

(Following summary information is given:  $\bar{x} = 441.67$ ;  $\bar{y} = 19.33$

$s_{xx} = 703315.67$ ;  $s_{xy} = 17125.11$ )

1. Make a scatterplot of x and y. **(1 mark)**

2. Estimate the coefficients of the linear regression equation  $Y = b_0 + b_1 X + e$ , using least squares method.

**(3 marks)**

3. Obtain the estimated linear regression equation.

**(1 mark)**

4. By how many rupees does the license fee increase for every increase of one billion rupees in total assets?

**(1 mark)**

5. Obtain the fitted values of license fees for all the observed  $x$  values, and a predicted value for a new bank with 250 billion rupees worth assets. **(2marks)**

Comment on the fit in relation to the scatterplot and on the prediction. What's your opinion? (You must justify)  
**(2 marks)**

**(Total 10 marks)**