# UNIVERSITY OF MORATUWA

Faculty of Engineering

Department of Mathematics

B.Sc. Engineering

Semester 3 Examination (2020 Batch)

## MA3014 – Applied Statistics

Time allowed: 2 hours                                      January 2023

**INSTRUCTIONS TO CANDIDATES:**

- This paper contains 4 questions on 4 pages.

- Answer **ALL** questions.

- The total maximum mark attainable is 100. The marks assigned for each question are indicated in square brackets.

- This examination accounts for 70% of the module assessment.

- This is a closed book examination.

- Assume reasonable meanings for any notation used and not defined in the examination paper. Clearly state any such assumptions made on the answer script.

- If you have any doubt as to the interpretation of the wording of a question, make your own decision, but clearly state it on the answer script.

- Electronic communication devices are not permitted.

- All examinations are conducted under the rules and regulations of the university.

**ADDITIONAL MATERIAL:**

- Statistical Tables

## Question 01

A) The actual proportion of men who favour a certain tax proposal is 0.40 and the corresponding proportion for women is 0.25. $n_1 = 500$ men and $n_2 = 400$ women are interviewed at random, and their individual responses are looked upon as the values of independent random variables having Bernoulli distributions with the respective parameters $p_1 = 0.40$ and $p_2 = 0.25$.

    i.    Specify the approximate sampling distributions of the sample proportions of men and women who favour the tax proposal.     *[06 Marks]*

    ii.    Find the 95% confidence interval of the difference between the two population proportions of favorable responses.     *[06 Marks]*

    iii.    Interpret the confidence interval found in part A) ii.     *[03 Marks]*

B) Consider taking samples of size 250 each from men and women after three months from the tax proposals are implemented to check whether the proportions still maintain a gap of at least 0.15 between them. The resulted proportions are 0.45 and 0.35, respectively for men and women. Using a statistical hypothesis test check whether the claim on the gap is true at 5% significance level.     *[10 Marks]*

## Question 02

A)

    i.    A hard disk manufacturing process deposits a thin coating of magnetic material on a plastic substrate. An empirical measurement has shown that the process introduces 20 defects per $15000\text{cm}^2$. A disk pack contains $800\text{cm}^2$ of this material. A disk pack with one defect is marketable because the disk controller can compensate for one error by using a spare location. Find the percentage of marketable disk packs.     *[05 Marks]*

    ii.    Suppose that a central processor receives interrupts under circumstances that justify a Poisson distribution. On the average, the processor receives 150 interrupts in a 10 millisecond period. What is the average and variance of the time between interrupts?

    *[05 Marks]*

B)    i.    The random variable Y has a normal distribution with an unknown population mean $\mu_Y$. A sample of size 20 yields $\bar{y} = 13.1$ and $s^2 = 140.42$. Taking this value as an estimate, compute the 90% confidence interval for population variance $\sigma_Y^2$.

    *[07 Marks]*

    ii.    Suppose the least square regression line $y = 0.208 + 0.026x$ is fitted for data collected on Y in part B) i. using a predictor variable X. If the Regression Mean Square of Y is

estimated as 120.97 and the Error Mean Square is 19.45. Test whether the above regression model fits the data well at 5% significance level. *[08 Marks]*

## Question 03

A) Under a special circumstance, state universities in a certain country decided to switch to online delivery mode from the existing physical mode. There was a claim that Engineering students of Intake2020 of a certain university preferred online education. Assuming that this preference will be maintained among the future intakes too, the university decided to switch to the online mode of delivering lectures in the Faculty of Engineering. But to be certain about the fact, an administrative officer decided to conduct a cross sectional study among students about their opinion on continuing online mode, when there are three batches who have the experience on the online delivery.

   i. Briefly describe an applicable sampling plan to select 350 students representing 3 intakes and 5 disciplines. *[05 Marks]*

Table 2.1 reveals some data obtained through this study.

**Table 2.1**

| | | Engineering field of study | | | | |
|---|---|---|---|---|---|---|
| | | A | B | C | D | E |
| Preferred mode of delivery | Online | 45 | 21 | 10 | 13 | 33 |
| | Physical | 25 | 23 | 23 | 22 | 12 |

   ii. Based on the data in Table 2.1, decide whether there is any significant association between the preference of students over the mode of delivery and the field of engineering studies. Use a suitable statistical testing procedure at 5% significance level to support your decision with scientific evidence. *[10 Marks]*

B) Suppose you found using the same approach as in part A) ii. that the students' SGPA (converted to 5 categories) of the last semester (done in online mode) is independent from their preference on mode of delivery, at 5% significance level. Further, you are supposed to compare the SGPA of students between the two groups of students (prefer Online vs prefer physical). Briefly describe the steps that you would follow sequentially in your analysis of checking whether there is no difference between the two groups.

(Your answer must be comprised of a description of data you wish to take, definitions of variables and parameters, any assumptions underlying the analysis, theoretical approach in statistics including any test statistic and distributional assumptions) *[10 Marks]*

## Question 04

A) An experiment is carried out to compare the effects of 5 types of stress releasing activities on IT professionals while they are engaged in regular projects. It was known that their performance indicator (a score) also depends on the type of the project, and on the time of the day. Assume there is a washout period so the previous activity will not affect the future results. Thus, the experiment is designed to obtain the performance score after the selected people are randomly allocated with the 5 types of stress releasing activities (D1, D2, D3, D4, D5) as in Table 4.1.

**Table 4.1**

| Type of the project | Time interval of the day | | | | |
|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| 1 | **D2**  38 | **D3**  39 | **D4**  45 | **D5**  41 | **D1**  40 |
| 2 | **D4**  32 | **D5**  37 | **D1**  38 | **D2**  30 | **D3**  35 |
| 3 | **D1**  35 | **D2**  36 | **D3**  37 | **D4**  32 | **D5**  33 |
| 4 | **D5**  33 | **D1**  30 | **D2**  35 | **D3**  33 | **D4**  30 |
| 5 | **D3**  36 | **D4**  38 | **D5**  39 | **D1**  35 | **D2**  41 |

i.

    a. How many treatments are there, and what are they? *[03 Marks]*

    b. How many replicates are there? *[02 Marks]*

ii. Write the statistical model applicable to describe the performance score in the above design of experiment. Clearly define all the parameters involved in the model.

*[05 Marks]*

iii. Test the hypothesis that there is no difference in performance of IT professionals when undergoing the different types of stress release activities, at 5% level of significance. *[08 Marks]*

B) The best fitted multiple linear regression model for a data set is given in equation 4.1.

$$\hat{y}_{ij} = 7.38 + 0.04x_1 + 0.07x_2 + 0.30x_3 - - - - \textbf{Equation 4.1}$$

i. Write two assumptions underpinning the above model. *[02 Marks]*

ii. Interpret the regression parameter of $X_3$. *[02 Marks]*

iii. The estimated $R^2$ value is 0.5582. What can you say about the model using this figure? *[03 Marks]*

**End of the paper.**