THE OPEN UNIVERSITY OF SRI LANKA

B.Sc. /B.Ed. Degree Programme, Continuing Education Programme

APPLIED MATHEMATICS-LEVEL 05

ADU5318 – Bio Statistics

222

FINAL EXAMINATION 2023/2024

Duration: Two Hours.

Date: 22.03.2024	**Time: 2.00p.m- 4.00 p.m.**

Instructions:

- **This question paper consists of 06 questions. Answer only four questions.**
- **Statistical Tables are provided. When reading values, where applicable, you may use the closest degrees of freedom given in the table.**
- **In all tests, use the significance level as 0.05.**
- **If the random number table is to be used, you are required to clearly indicate how to read the values and as an illustration present three values that you read.**
- **Non-programmable calculators are permitted.**

1. A manufacturer claims that the mean lifetime of car batteries produced by his company exceeds 4 years. The sample mean and the standard deviation of the lifetimes (years) of 20 randomly selected car batteries were 3.2 and 0.8 respectively.

   i) Clearly describing the notation you use, write down the appropriate null hypothesis and the alternative hypothesis to examine the validity of the manufacturer's claim.

   ii) Suppose that the lifetimes of the said batteries follow a normal distribution. Using a 5% significance level, test the validity of the manufacturer's claim. Clearly state the findings.

   iii) Explain each of the following terms in relation to this study.

      a) Type I error

      b) Critical region

2. A study is to be designed to compare the effects of three drugs (say D1, D2, D3) on the blood glucose levels of rats. The researcher has access to 60 rats of age less than three months and 90 rats of age around eight months. Rats in each group had similar blood glucose levels at the start of the study. The researcher suspects possible differences in reactions to the drugs depending on the age of the rat. Two methods of drug administration, which are: drug given in the morning (say M1) and drug given in the night (say M2) are to be tested. Changes in the blood glucose levels are to be measured after one month of drug administration. Suppose there are no limitations on the drug availability for this study.

Suppose that the researcher seeks your advice to design this study.

i) Clearly describe how you advise the researcher.

ii) Explain each of the following terms, in relation to this study
   a) Replicate
   b) Confounding
   c) Interaction

iii) Briefly describe two advantages of using replicates in a research study.

3. A production process is considered to be running at an acceptable level, only when the defective percentage among items produced is maintained below 3%. The management regularly monitors the production process by examining samples of items produced and counting the number of defectives among sampled items. The production process has two shifts, a day shift and a night shift. The main objectives of the management are: to find out whether the production process is running at an acceptable level, whether there are differences between day shift and the night shift with regards to defective percentages and how the defective percentages in each shift has changed over time. The following table presents the data collected from the two shifts over a selected time period.

Description		Jan 2022	May 2022	Aug. 2022	Jan 2023	June 2023	Oct 2023	Jan 2024	Feb 2024
Day shift	No. of items sampled	140	152	160	140	150	160	140	150
	No. of defectives	4	5	6	3	5	5	5	6
Night shift	No. of items sampled	70	60	45	72	74	44	56	62
	No. of defectives	3	2	2	3	3	1	1	1

i) Construct a suitable graph to address the objectives of the management.

ii) Clearly state the findings from the graph constructed in part (i).

iii) Suppose the management is able to correctly identify good items but may misclassify defective items as good items. Will this misclassification increase the sampling error, reduce the sampling error or will have no effect on the sampling error? Give reasons for your answer.

iv) Does the graph constructed in part (i) support the statement that, for the purpose of estimating the defective percentage of the production process, the population is a homogeneous population? Give reasons for your answer.

4. The following cumulative frequency table was constructed from the reaction times (measured to the nearest minute) of a collection of chemical samples used in a research study.

Time (in minutes)	Cumulative frequency
10 - 14	4
15 - 19	6
20 - 24	8
25 - 29	10
30 - 34	12
35 - 39	18
40 - 44	30

i) How many chemical samples were used for measuring the reaction times summarised in the table?

ii) Calculate the proportion of samples that had reaction times
   a) less than 20 minutes
   b) in the range from 15 to 29 minutes.

iii) Calculate
   a) Sample median
   b) First quartile
   c) Inter-quartile range

iv) State whether the values of each of the three measures calculated in part (iii), would change or remain unchanged, if the data values used to construct the given table were corrected for the errors, as described in each of the following situations. In each case, provide reasons for your answer.
   a) A measurement error was corrected by adding five minutes to each data value.
   b) A data value that had been incorrectly recorded as 14 was corrected as 24.

5. The management of a hospital wants to assess the satisfaction of persons regarding the facilities available. The data are to be collected from persons attending the eye clinic, diabetic clinic and the wellness clinic, on a randomly chosen day. Around 40% of persons attending each clinic are females and the researcher suspects gender

differences in their opinion on facilities. On a single day, the hospital expects around 300, 600 and 100 persons attending the eye clinic, diabetic clinic and the wellness clinic respectively. Resources are available to sample 200 persons for data collection.

i) Suggest a suitable sampling design. You need to clearly explain how you sample the persons.

ii) State the data type as nominal, ordinal, interval or ratio in each of the following variables:

   a) Type of clinic recorded using codes: 1: Eye clinic; 2: Diabetic clinic: 3: Wellness clinic
   b) Gender of the person attending the clinic
   c) Age of the person recorded as 1: less than 20 years; 2: 21 to 30 years; 3: 31 to 40 years; 4: above 40 years
   d) Waiting time (in hours) of the person for medical advice
   e) Total number of previous visits to the clinic

iii) State whether each of the following statements is true or false. In each case, give reasons for your answer.

   a) Bias can be reduced by increasing the sample size.
   b) If the researcher had sampled all the persons who attended these clinics on that day, non-sampling errors cannot occur.

6. The following table summarises the yield of paddy per acre, measured to the nearest bushel, of a group of farmers randomly selected for a research study.

Yield (bushels per acre)	Number of farmers
15 – 24	10
25 – 34	12
35 - 44	30
45 – 54	34
55 – 64	14

i) Calculate the sample mean and explain what it measures in relation to this study.

ii) Calculate the sample median.

iii) Based on the measures calculated in parts (i) and (ii), what can you say about the distribution of the data?

iv) Out of the two measures calculated in part (i) and part (ii), which measure would you recommend to estimate the population mean of yield per acre. Give reasons for your choice.

v) In relation to this study, explain what is meant by the dispersion of the data and propose a suitable measure. **You need not calculate the value of the proposed measure.**

vi) Suggest a suitable graph that can be used to find the percentiles of the data. **You need not construct the graph.**

******* Copyrights reserved ********