NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

SORS2110

FACULTY OF APPLIED SCIENCES

DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH

SORS2110: INTRODUCTION TO APPLIED STATISTICS

AUGUST 2024 SUPPLEMENTARY EXAMINATION

Time : 3 hours

Candidates should attempt **ALL** questions from Section A and **ANY THREE** questions from Section B. Each question should start on a fresh page. Statistical tables, graph paper required

SECTION A: Answer all questions in this section (40 marks).

A1. Suppose you are attending a conference organized by the Faculty of Applied Sciences and it is presumed that all student scientist at the university attend the conference. It is perceived that some students think that there are no differences between scientists and engineers in their professional conduct. You thought of validating this perception. As a result, you collected a sample of student scientists attending the conference and administered a questionnaire. Conclusions on this perception should be drawn before the conference ends because it would be difficult to meet the student scientists in the Faculty as group like that again.

(a) Which sampling method would you use? Explain your answer. [1,2]

(b) How would you administer the questionnaire? Justify your answer. [1,2]

A2. A machine fills packets with a chlorine chemical which is used to treat swimming pools. A packet is expected to have a weight of 40 grams. In a quality control inspection, a sample of 36 packets was taken and they were found to have an average of 42.4 grams with a standard deviation of 6 grams. Test whether this machine is filling the packets to the expected weight. [5]

A3. (a) The average rainfall, recorded to the nearest hundredth of a centimetre, in Bulawayo, for the month of January is 9.22 centimetres. Assuming that rainfall in normally distributed with a standard deviation of 2.83 centimetres, find the probability that next January Bulawayo receives:

(i) At most 1.84 centimetres of rain [2]

(ii) More than 5 cm but not over 7 cm [3]

(iii) More than 13.8 cm [3]

(b) State the strengths and weaknesses of each of the following measures of central tendency in statistics:

(i) The Arithmetic Mean [4]

(ii) The Median. [4]

A4. The following data constitute the final examination marks (out of 100%) for a class which did a Statistics course:
28 49 31 36 41 26 52 39 46 40 37 58 47 43 50 34 54 28 48 40

(a) Construct a stem and leaf plot of the examination scores and comment on the distribution of the examination scores [4]

(b) What is the median examination score? [2]

(c) What is the modal examination score? [2]

(d) What is the average examination score? [3]

(e) Explain why would you think that the average is the best measure to describe the examination scores [3]

SECTION B: Answer any three questions in this section (60 marks).

B5. An experiment was performed in order to investigate the effect of four different fuels and three different types of launchers on the range of a certain military rocket and the following data were obtained, (range in km):

	Fuel 1	Fuel 2	Fuel 3	Fuel 4
Launcher I	45.9	57.6	52.2	41.7
Launcher II	46.0	51.0	50.1	38.8
Launcher III	45.7	56.9	55.3	48.1

(a) Test at the 5% level of significance whether there are significant differences among the fuels and also among the three launchers [10]

(b) Compute the standard error for comparing any two fuel means. Hence compare, at the 5% level of significance, each of **Fuels 2, 3**, and 4 with the standard **Fuel 1** [10]

B6. An Influenza vaccine was given to 2000 men, and 20 of these showed an allergic reaction. Of the 1750 women given the same drug, 50 showed an allergic reaction

	Gender		
Allergic reaction	Male	Female	Total
Present			
Absent			
Total	2 000	1 750	

(a) Complete table above and then test at the 5% level of significance whether an association exists between gender and an allergic reaction to the drug [10]

(b) In a photographic process, the developing time of prints may be looked upon as a random variable having a normal distribution with a mean of 15.40 seconds and a standard deviation of 0.48 seconds. Find the probabilities that the time it takes to develop one of the prints will be:

(i) At least 16.00 seconds [3]

(ii) At most 14.20 seconds [3]

(iii) Anywhere from 15.00 to 15.80 seconds (*Round your computed z values to two decimal places*) [4]

B7. A test was performed in order to determine the relationship between the chemical content of a particular constituent, y, *in grams per litre*, in solution and crystallization temperature,x, *in degrees Kelvin.* The results were as follows:

Chemical content,y	0.3	0.4	1.2	2.3	3.1	4.2	5.3
Crystallization temperature,x	3.2	2.4	4.3	5.4	6.6	7.8	8.8

(a) Draw a scatter diagram of y against x. Fit the least squares regression line $y = \beta_0 + \beta_1 x + \epsilon$ and draw your fitted line on your scatter diagram [10]

(b) Carry out an analysis of variance to test at the 0.05 level of significance whether the slope is significantly different from zero. Compute the coefficient of determination, R^2, and interpret it [8]

(c) Predict the chemical content, **y**, at the crystallization temperature of 3.6 degrees Kelvin. Compute the standard error of the predicted value and hence find the 95% confidence interval of this prediction. [7]

B8. (a) The following are the systolic blood pressure readings (*in mmHg*) taken on both arms of each member of a sample of ten students from the 2022 SORS1201 Conventional Class with fifty six students in total:

Student	1	2	3	4	5	6	7	8	9	10
Left	128	142	112	160	146	123	111	114	120	117
Right	121	146	105	141	138	121	124	112	122	123

Test at the 5% level of significance whether there is any significant difference in systolic blood pressure readings between the left and right arms. [10]

(b) A manufacturer of lead-covered components suspects that the right-hand components are receiving on average the same thickness of lead covering as the left-hand ones. He tells you that two random samples of size eight each gave the following results of mean thickness (cm x 10^{-3}):

Left-hand compo	366	362	376	399	355	332	382	340
Right-hand compo	338	346	379	363	312	321	360	301

Carry out an appropriate t-test to determine whether the manufacturers suspicion is valid at the 5% level of significance. [10]

LIST OF SELECTED FORMULAE

1. • $s^2 = \dfrac{\sum x_i^2 - \frac{(\sum x_i)^2}{n}}{n-1}$ and $s^2 = \dfrac{\sum f_i x_i^2 - \frac{(\sum f_i x_i)^2}{n}}{n-1}$

2. • $M_o = L_{mo} + \dfrac{c\left(f_{mo} - f_{mo-1}\right)}{2f_{mo} - f_{mo-1} - f_{mo+1}}$

   • $M_e = L_{me} + \dfrac{c\left(\frac{n}{2} - F_{me-1}\right)}{f_{me}}$

3. • $z_{calc} = \dfrac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$

   • $t_{calc} = \dfrac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$

4. • $\chi^2 = \sum \dfrac{(O_{ij} - E_{ij})^2}{E_{ij}}$; where, Expected Frequency $= \dfrac{\text{Row TotalColumn Total}}{\text{Grand Total}}$

5. • $z_{calc} = \dfrac{\bar{x_1} - \bar{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$

   • $t_{calc} = \dfrac{\bar{x_1} - \bar{x_2}}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ where $S_p^2 = \dfrac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$

   • $t_{calc} = \dfrac{\bar{d}}{\frac{s_d}{\sqrt{n_d}}}$

6. • $\hat{\beta}_1 = \dfrac{n\sum x_i y_i - \sum x_i \sum y_i}{n\sum x_i^2 - \left(\sum x_i\right)^2}$

   • $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

   • $r = \dfrac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{\left[n\sum x_i^2 - \left(\sum x_i\right)^2\right]\left[n\sum y_i^2 - \left(\sum y_i\right)^2\right]}}$

   • $TotalSS = \sum y^2 - \frac{(\sum y)^2}{n}, \quad RegressionSS = \hat{\beta}_1\left(\sum xy - \frac{\sum x \sum y}{n}\right)$

7 • $TotalSS = \sum_{ij} y_{ij}^2 - \dfrac{\left(\sum_{ij} y_{ij}\right)^2}{N}$

   • $TreatmentSS = \sum_i \dfrac{T_{i.}^2}{r_i} - CF, \quad BlockSS = \sum_j \dfrac{B_{.j}^2}{r_j} - CF$

END OF QUESTION PAPER