NATIONAL UNIVERSITY OF SCIENCE AND TECHNOLOGY

FACULTY OF APPLIED SCIENCE

DEPARTMENT OF STATISTICS AND OPERATIONS RESEARCH

SORS 1103 : INTRODUCTION TO STATISTICS

BSc Operations Research & Statistics, Business Analytics: Part I

December 2024 EXAMINATION

Time : 3 hours

Total Marks: 100

Candidates may attempt **ALL** Questions in Section **A** and at most **THREE** Questions in Section **B**. Show all working where necessary. You may use a calculator, Graph paper and Statistical Tables will be provided.

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

A1. (a) Define Statistics. [2]

(b) What is the difference between quantitative and qualitative data? [4]

(c) What is an outlier? State two reasons for the occurrence of outliers. [4]

A2. (a) The heights obtained on 12 learners are listed below .
1.03 1.32 1.05 1.00 1.13 1.06 1.12 1.03 1.10 1.24 1.42 1.37 .
Determine

(i) the mean. [2]

(ii) the standard deviation. [3]

(iii) a 99% confidence interval for the mean height. [5]

A3. (a) Distinguish a Type I error from Type II error. [4]

(b) Due to differences in the environment, the masses of certain species of small animals are believed to be greater in region A than in region B. It is known that the masses in both regions are Normally distributed with masses in region A and region B having a standard deviation of 0.04kg and 0.09kg respectively. A sample of 60 animals was taken from region A and 50 animals from region B, the samples had mean masses 3.03kg and 3.00kg respectively. Does this provide evidence at 1% level of significance that the animal species in region A have greater mass than those in region B? [6]

A4. Table 1 is a partially completed ANOVA table for a completely randomised design.

Table 1: Incomplete Summary ANOVA table

Source	df	SS	MS	F
Treatments	**	2709.20	**	35.81
Error	36	**	**	
Total	39	3617.06		

(a) Copy and complete the ANOVA table. [2]

(b) State the number of treatments involved in the experiment. [1]

(c) Do the data provide sufficient evidence to indicate a difference among the treatment means? Test by using $\alpha = 0, 05$. [7]

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

B5. (a) State any 2 characteristics of the F-distribution and state any 2 assumptions of the One-way ANOVA. [4]

(b) A large company buys thousands of light bulbs every year. It is currently considering three brands to choose from. Before the company decides it wants to investigate if the mean life times of the bulbs are the same. The company's research department randomly selected ten bulbs of each type and tested them. The results of number of hours that each bulb lasted before burning out are as follows in Table 2

Table 2: Time taken in thousand of hours for bulbs to burn out

Type A	75	80	75	77	77	76	81	77	74	82
Type B	79	69	76	76	.74	81	77	82	81	84
Type C	773	794	733	740	780	801	794	719	766	743

Test at 1% level of significance that the mean burning times for all bulbs is the same. [16]

B6. A random sample of 207 people was used to collect data on the amount of time spent at the gym per day . The results are displayed in Table 3

Table 3: Time spent at the gym

Time spent (t minutes)	0-15	15-30	30-45	45-60	60-75	75-90
Number of people	21	52	45	41	37	11

(i) State the modal group . [1]

(ii) Calculate the approximate values of the mean and standard deviation of the time spent per day at the gym. [7]

(iii) Estimate the 99% mean confidence interval for the time spent at the gym. [4]

(iv) Draw a cumulative frequency curve and use it to estimate the median. [8]

B7. Table 4 shows information on the price of a magazine and the percentage of the magazine space that contains advertisements.

Table 4: Price of a magazine and percentage space that contains ads

Percentage	37	43	58	49	70	28	65	32	27	14	20	23
Price($)	5.50	6.95	4.95	5.75	3.95	8.25	5.50	6.75	6.00	7.75	6.5	7.50

(i) Construct a scatter plot for these data. [4]

(ii) Is it suitable to fit a linear regression model and why? [2]

(iii) Find the predictive regression equation of price on the percentage containing ads. [6]

(iv) Interpret the values of a and b in part (iii). [2]

(v) Calculate the correlation coefficient for the data and interpret it. [3]

(vi) Predict the price of a magazine with 50% of its space containing ads. [3]

B8. (a) State the parameter for Chi-square distribution. [1]

(b) Distinguish between a goodness of fit test and a test for independence. [4]

(c) A forestry official is comparing the causes of fires in three towns E, F and G. The following table shows the causes of fire for 128 randomly selected fires in these 3 towns.

Table 5: Causes of fire

	Arson	Accident	Lightining	Unknown
Town E	16	9	6	10
Town F	8	14	15	9
Town G	5	8	9	19

(i) State the null and alternative hypotheses. [2]

(ii) At the 5% level of significance, can you say that the causes of fire are related to the towns encountering the fire ? [13]

LIST OF SELECTED FORMULAE

Population Proportion	Two Population Proportions	Population Mean
Parameter p	Parameter p_1-p_2	Parameter μ
Statistic $\hat{p}$	Statistic $\hat{p}_1$-$\hat{p}_2$	Statistic $\hat{\mu}$
Standard Error $s.e(\hat{p}) = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	Standard Error $s.e(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n_1} + \frac{\hat{p}_2(1-\hat{p}_2)}{n_2}}$	Standard Error $s.e(\bar{x}) = \frac{s}{\sqrt{n}}$
Confidence Interval	Confidence Interval	Confidence Interval
$\hat{p} \pm z * s.e(\hat{p})$	$(\hat{p_1} - \hat{p_2}) \pm z * s.e(\hat{p_1} - \hat{p_2})$	$\bar{x} \pm t^* s.e(\bar{x})$ df=$n-1$ Paired Confidence Interval $\bar{d} \pm t^* s.e(\bar{d})$ df=$n-1$
Large-Sample z-Test $z_0 = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	Large-Sample z-Test $z_0 = \frac{\hat{p_1}-\hat{p_2}}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$ where $\hat{p} = \frac{x_1 + x_2}{n_1 + n_2}$	One-Sample t-test $t_0 = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$ df=$n-1$ paired t-Test $t_0 = \frac{\bar{d}}{\frac{s_d}{\sqrt{n}}}$ df=$n-1$

Table 6: Two Population Means

General	Pooled
Parameter $\mu_1 - \mu_2$	Parameter $\mu_1 - \mu_2$
Statistic $\mu_1 - \mu_2$	Statistic $\mu_1 - \mu_2$
Standard Error $s.e(\bar{x}_1 - \bar{x}_2) = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$	Standard Error pooled s.e$(\bar{x}_1 - \bar{x}_2) = s_p\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ where $s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}$
Confidence Interval $(\bar{x}_1 - \bar{x}_2) \pm t * s.e(\bar{x}_1 - \bar{x}_2)$ df=min$(n_1 - 1, n_2 - 1)$	Confidence Interval $(\bar{x}_1 - \bar{x}_2) \pm t *$ (pooled s.e$(\bar{x}_1 - \bar{x}_2)$) df=$n_1 + n_2 - 2$
Two-Sample t-Test $t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ df=min$(n_1 - 1, n_2 - 1)$	Pooled Two-Sample t-Test $t_0 = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{s_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ df=$n_1 + n_2 - 2$ df=min$(n_1 - 1, n_2 - 1)$

END OF QUESTION PAPER

Table 7: One-Way ANOVA

SS treatment(Groups)=SST=$\sum_{i=1}^{k} n_i(\bar{x}_i - \bar{x})^2$	MST=$\frac{SST}{k-1}$
SS Error=SSE= $\sum_{i=1}^{k}(n_i - 1)s_i^2$	MS Error =MSE=$\frac{SSE}{n-k}$
Total SS =$\sum x_{ij}^2 - \frac{(\text{grand total})^2}{n}$	F=$\frac{MST}{MSE}$
Note: n=total number of observations	

Table 8: Regression

Standard Error of the Sample Slope $s_{b_1} = \frac{S_e}{\sqrt{S_{XX}}} = \frac{S_e}{\sqrt{\sum(x-\bar{x})^2}}$	Confidence Interval for β_1 $b_1 \pm t * s_{b_1}$ df=$n-2$
Estimate of σ $S_e = \sqrt{MSE} = \sqrt{\frac{SSE}{n-2}}$ where SSE=$\sum(y-\hat{y})^2 = \sum e^2$	t-test for β_1 $t_0 = \frac{b_1}{s_{b_1}}$ $df = n-2$ or F=$\frac{MSREG}{MSE}$ $df = 1, n-2$
Confidence Interval for Mean Response $\hat{y} \pm t_{\alpha/2} \cdot S_e\sqrt{\frac{1}{n} + \frac{(x*-\bar{x})^2}{\sum(x_i-\bar{x})^2}}$	Prediction Interval for an Individual Response $\hat{y} \pm t_{\alpha/2} \cdot S_e\sqrt{1 + \frac{1}{n} + \frac{(x*-\bar{x})^2}{\sum(x_i-\bar{x})^2}}$

Table 9: Chi-Square Tests

Test of Independence & Test of Homogeneity	Test for Goodness of Fit
Expected Count E=$\frac{\text{row total}\times\text{column total}}{\text{grand total}}$	Expected Count E=np_i
Test Statistic $\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$ df=(r-1)(c-1)	Test Statistic $\chi^2 = \sum \frac{(\text{Observed} - \text{Expected})^2}{\text{Expected}}$ df=k-1