320

(University of Choice)

MASINDE MULIRO UNIVERSITY OF SCIENCE AND TECHNOLOGY
(MMUST)
MAIN CAMPUS

UNIVERSITY EXAMINATIONS 2021/2022 ACADEMIC YEAR

SECOND YEAR SECOND SEMESTER EXAMINATIONS

FOR BACHELOR

OF

SCIENCE IN INFORMATION TECHNOLOGY

COURSE CODE: BIT220

COURSE TITLE: DATA ANALYSIS TECHNIQUES

DATE: 19/04/2022 **TIME:** 8:00-10:00 AM

INSTRUCTIONS TO CANDIDATES

Answer **Questions ONE** and **ANY OTHER TWO**

TIME: 2 Hours

MMUST observes ZERO tolerance to examination cheating

QUESTION #1 [COMPULSORY- 30 MARKS]

West kenya Sugar Company is primarily engaged in the manufacture and sale of sugar. The company grows some sugar cane; its own estates provide up to 7% of its annual output. Its primary source of sugarcane is over 50,000 registered "out growers" with over 400 square kilometres (99,000 acres) under cultivation. It has also piloted the production of a hybrid high-yielding palm oil variety in areas previously thought too cool for commercial cultivation, in collaboration with the UN Food and Agriculture Organization (FAO). In addition to sugar, the company co-generates 34 Megawatts of electricity. Some of the electric power is used internally and surplus is sold into the nation electricity grid. The company also manufactures 24 million liters of ethanol annually and 20 million liters of distilled water. In the last few years of its establishment, it has experienced an acute labour turnover, a trend that has worried the management. You have been contracted by the COMPANY to conduct a research to establish the causes of this problem, Using this hypothetical information, answer the following questions:

a) Formulate the research topic [3 marks]
b) Identify one independent and two independent variables for this study [3 marks]
c) State any four specific objectives that would guide the study [3 marks]
d) Enumerate the four research questions that will be used in the study [4 marks]
e) Formulate any four research hypotheses that will guide the study [4 marks]
f) Which instruments of data collection will you use during the conduction of research?[3 marks]
g) Distinguish between the implications skewedness and kurtosis. Use diagrams to illustrate their various types [4 Marks]
h) Explain three measures of central tendency [3 Marks]
i) State and explain using an example three categories of descriptive statistics that you would use to analyze your data [3Marks]

QUESTION #2 [20 MARKS]

Interpret the output from the regression including answering the following questions:

a) Which independent variables are significant and what is their relationship with the dependent variable? (6MARKS)

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-7.165	2.107		-3.400	.002
	Gestational age at birth (weeks)	.313	.053	.623	5.926	.000

Smoker	-.665	.268	-.253	-2.485	.017
Mothers pre-pregnancy weight (lbs)	.020	.009	.237	2.261	.030

a. Dependent Variable: Weight of baby at birth (lbs)

b) What is the equation of the model? (3 MARKS)

c) How reliable is the model? (3MARKS)

a) Briefly explain the relevance of the Chi-Square Tests [4 Marks]

b) You have been asked to examine the relationship between the Gender of the student and the number of other colleges the student has visited before settling at MMUST. The Data is as below:

ID	Gender 1=Female, 2=Male	Colleges visited before
22657781	1	3
22667782	2	4
23678883	2	7
11178984	1	5
11767115	1	2
13131816	2	1
11314147	2	6
12131418	1	4
14151619	1	2

Explain the steps you are likely to follow in order to calculate correlation coefficient in SPSS. [4 Marks]

QUESTION #3 [20 MARKS]

A researcher is interested to see the effect of studying (x_1) or sleeping (x2) on a test score(y). In the example, "test score" is the dependent variable. "Studying" or "sleeping" is the independent

variable because these factors impact how much a student scores on the test. To answer this question, he collects the data from a one stream school as shown in the table below.

School	Studying (x1)	Sleeping (x2)	Test score (y)
K	5	3	7.312
M	8	2	8.478
L	7	3	6.133

(a) State the null hypothesis for this study [2 Marks]

(b) Using SPSS explain how you will be able to compute the mean and standard deviation of Studying (x1), Sleeping (x2) and Test score (y) [5 Marks]

(c) Explain how you will use SPSS to compute Pearson's correlation coefficient r for

1. Studying and Test score. [4 Marks]
2. Sleeping and Test score [4 Marks]

(d) Explain how you will use SPSS to compute a linear regression analysis and produce the linear regression equation of x1 x2 (independent variables) and y (dependent variable) in the form:

$$y = a + bx_1 + cx_2$$ [5 Marks]

QUESTION #4 [20 MARKS]

a) Explain briefly why ANOVA is called Analysis of variance instead of Analysis of means. (4marks)

b) What are the assumptions for ANOVA and how can they be tested? (3marks)

d) In SPSS, how can one combine values of a variable into smaller number of categories? [3mks]

e) Discuss if there is a "correct" order or steps for a statistical analysis? [3mks]

f) For each of the following research situations identify which statistical technique could be used. [7mks]

i. Ann is interested in exploring the possibility of gender differences in levels of perceived stress.
ii. Ann would also like to explore the relationship between optimism and perceived stress. She suspects that higher levels of optimism would be associated with lower levels of perceived stress.

iii. Bill is interested in exploring the effect of both sex and age group on self-esteem scores. He is interested in the effect of each variable individually, and any interaction that may exist.

iv. Celia would like to know which is a better predictor of negative affect: optimism or self-esteem.

v. If Celia were also concerned that age may be a confounding variable, how would she go about controlling for this variable in the analyses?

vi. David is interested in the question: Are younger people (18-29yrs) more likely to be smokers than older people (30-44yrs or 45+yrs)?

vii. Ellie conducts a study to find out if there is a significant change in depression levels across three time periods (prior to an intervention, after the intervention and at a three-month follow-up).

QUESTION #5 [20 MARKS]

Consider sample questionnaire shown below

Demographic Questionnaire

Here's a template for surveying customers to learn more about their demographic background. You could substantiate the analysis of this questionnaire by corroborating the data with other information from your web analytics, internal customer data, and industry data.

1. How would you describe your employment status?
   a. Employed full-time
   b. Employed part-time
   c. Freelance/contract employee
   d. Self-employed
   e. Unemployed

2. How many employees work at your company?
   a. 0-20
   b. 21-50
   c. 51-100
   d. 101-1,000
   e. 1,001+

3. How would you classify your role?
   a. Individual Contributor
   b. Manager
   c. Director
   d. VP
   e. Executive

4. How would you classify your industry?

a. Technology/software
b. Hospitality/dining
c. Services
d. Consulting
e. Entertainment

a) Explain how you are going to perform coding of the various variables for the Questionnaire above [8 Marks]

b) **C**ross tabulation is a way of examining the relationship between two variables. State the procedure likely to be followed when you want to carry out cross tabulation with SPSS [5 Marks]

c) Prepare a codebook from the questionnaire above (7marks)