TECHNICAL UNIVERSITY OF MOMBASA

FACULTY OF APPLIED AND HEALTH SCIENCES

MATHS/ PHYSICS DEPARTMENT

UNIVERSITY EXAMINATION FOR:

BACHELOR OF SCIENCE IN STATISTICS AND COMPUTER SCIENCE

AMA 4308: APPLIED STATISTICAL METHODS PAPER 1

END OF SEMESTER EXAMINATION

SERIES: MAY 2016

TIME: 2 HOURS

DATE: MAY 2016

Instructions to Candidates

You should have the following for this examination

-Answer Booklet, examination pass and student ID

This paper consists of FIVE questions. Attempt question ONE COMPULSORY and any other TWO from 2-5

Do not write on the question paper.

Question ONE (30 marks)

a. Define the following terms
   I. Regression (2 marks)
   II. Response surface methodology (2marks)
   III. Analysis of variance (2 marks)
   IV. P- value (2marks)
b. Outline five uses of regression models (5marks)
c. Prove that $\sum_{jk}(X_{jk}-\bar{x})^2 = \sum_{jk}(X_{jk}-\bar{x}_{j.})^2 + \sum_{jk}(X_{j.}-\bar{x})^2$ (4marks)
d. Distinguish between a probabilistic model and a deterministic model with aid of graphs.(4marks)
e. State one assumption of regression models (1 mark)
f. State four characteristics of a good experimental design (4 marks)
g. $y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 +, \dots, \beta_K X_K + \varepsilon$
   Identify the stated equation (1mark)
   From the equation, state the
   I. Independent variable (1 mark)
   II. Dependent variable (1mark)
   III. The stochastic component (1 mark)

Question TWO (20 marks)

The table below shows the yields in bushels per acre of a certain variety of wheat grown in a particular type of soil treated with chemicals A, B or C. Find ;

a. The mean yield for the different treatments (3 marks)
b. The grand mean for all treatments (2 marks)
c. The total variation (2 marks)
d. The variation between treatments (2 marks)
e. The variation within treatments (2 marks)
f. Test the null hypothesis of equal means
    I. At 0.05 significance level (3marks)
    II. At 0.01 significance level (2marks)
g. Represent the information in an ANOVA table (4 marks)

A	48	49	50	49
B	47	49	48	48
C	49	51	50	50

3	4	5	4
2	4	3	3
4	6	5	5

Question THREE

The table shows yields per acre of four different plant crops grown on lots treated with three different types of fertilizer. Using the long method and representing the information in an ANOVA table test at 0.01 level of significance whether;

a. There is a significant difference in the yield per acre due to fertilizers (9 marks)
b. There is a significant difference in yields per acre due to crops (9 marks)

	Crop 1	Crop II	Crop 111	Crop IV
Fertilizer A	4.5	6.4	7.2	6.7
Fertilizer B	8.8	7.8	9.6	7.0
Fertilizer C	5.9	6.8	5.7	5.2

c. Define the parameters of a two factor experiment (2 marks)

Question FOUR (20marks)

The yield of a batch process in the chemical industry is known to be approximately linearly related to temperature, at least over a limited range of temperatures. Two measurements of the yield are made at each of the eight temperatures within this range with the following results

Temp($c^{\cdot 0}$) x	180	190	200	210
Yield (tonnes) y	136.2 136.9	147.5 145.1	153.0 155.9	161.7 167.8
Temp($c^{\cdot 0}$) x	220	230	240	250
Yield (tonnes) y	176.6 164.4	194.2 183.0	194.3 175.5	196.5 219.3

$\sum x = 172 \quad \sum x^2 = 374000$

a. Plot the data on a scatter diagram (4 marks)
b. For each temperature, calculate the mean of the two yields (4 marks)
c. Calculate the equation of the regression line of this mean yield on the temperature(7 Marks)
d.
e. Predict from the regression line the yield of a batch at each of the following temperatures
    I. 175 (1mark)
    II. 185 (1mark)
    III. 300 (1mark)

Question FIVE (20 marks)

a. Define the following terminologies
    I. Polynomial regression (2marks)
    II. Ridge regression (2marks)
    III. Multiple linear regression (2marks)
    IV. Collinearity (2 marks)
b. Define the term adjusted R^2 and discuss how it is used to test for the fitness of a model (8 marks)
c. Prove that $\sum_{jk}(X_{jk} - \bar{x})^2 = \sum_{jk}(X_{jk} - \bar{x}_{j.})^2 + \sum_{jk}(X_{j.} - \bar{x})^2$ (4marks)