JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY

SCHOOL OF MATHEMATICS AND ACTUARIAL SCIENCE

UNIVERSITY EXAMINATION FOR DEGREE OF BACHELOR OF SCIENCE ACTUARIAL

3rd YEAR 2nd SEMESTER 2016/2017 ACADEMIC YEAR

MAIN REGULAR

SAC 306: STATISTICAL MODELLING 1

EXAM VENUE: **STREAM:**

DATE: EXAM SESSION:

TIME: 2.00 HOURS

Instructions:

1. **Answer question 1 (Compulsory) and ANY other 2 questions**
2. **Candidates are advised not to write on the question paper.**
3. **Candidates must hand in their answer booklets to the invigilator while in the examination room.**

Question One Compulsory (30 marks)

a) We're doing two-way ANOVA with four categories of Factor A, seven categories of Factor B, and two observations per group. The ANOVA table that we get is shown below, but with some items missing:

Source	*df*	*SS*	*MS*
Factor A	3	11.84	3.95
Factor B	6	6.78	1.13
Interaction	18	17.10	0.95
Error		16.76	
Total		52.48	

One of the missing items is MSE. Calculate it. (3mks)

b) We're doing one-way ANOVA with 30 total observations among all the groups combined. Our ANOVA table is shown below, but with some items missing:

Source	*df*	*SS*	*MS*
Group		9.99	
Error	25	23.56	0.94
Total		33.55	

One of the missing items is dfG. Calculate it. (3mks)

c) You're using two-way ANOVA to study how students' absences at your local high school are affected by two factors: class year (freshman, sophomore, etc.) and geographic location of the student's home. You have four categories of class year, seven categories of home location, and two observations per group. (All assumptions are met well enough to proceed.) You obtain the following ANOVA table:

Source	*df*	*SS*	*MS*
Class Year	3	26.20	3.32
Home Location	6	9.16	1.53
Interaction	18	20.98	1.17
Error	28	73.76	2.63
Total		130.10	

You first test for interaction and see no evidence of any, so you now want to test for main effects.

Calculate the test statistic for the main effect of home location. (5mks)

d) Briefly state four assumptions of ANCOVA (4mks)

e) Define the following terms (4mks)

i) ANCOVA

ii) Outlier

iii) Collinearity

iv) Muliticollinearity

f) Bernouilli's equation for continuous flow out of a tank relates the volumetric flow, Q, to the height of fluid in the tank, h:

$$Q = k \cdot h^a$$

Data for flow rate as a function of the liquid level in the tank is given in the table below.

h (cm)	Q (li/min)
5	0.2322
40	0.6450
70	0.8207
100	0.9856

(i) Estimate the values for k and a. (7mks)

(ii) Estimate k if a is assumed to be 0.5. (4mks)

Question Two (20 marks)

a) A firm wishes to compare four programs for training workers to perform a certain manual task. Twenty new employees are randomly assigned to the training programs, with 5 in each program. At the end of the training period, a test is conducted to see how quickly trainees can perform the task. The number of times the task is performed per minute is recorded for each trainee, with the following results:

Observation	*Program 1*	*Program 2*	*Program 3*	*Program 4*
1	9	10	12	9
2	12	6	14	8
3	14	9	11	11
4	11	9	13	7
5	13	10	11	8
$T_{Aj} = \Sigma y_{ij}$	59	44	61	43
$\hat{\mu}_j = T_{Aj}/N_j$	11.8	8.8	12.2	8.6

Estimate the treatment effects for the four programs. (4mks)

b) A firm wishes to compare four programs for training workers to perform a certain manual task. Twenty new employees are randomly assigned to the training programs, with 5 in each program. At the end of the training period, a test is conducted to see how quickly trainees can perform the task. The number of times the task is performed per minute is recorded for each trainee, with the following results:

Program 1: 9, 12, 14, 11, 13
Program 2: 10, 6, 9, 9, 10
Program 3: 12, 14, 11, 13, 11
Program 4: 9, 8, 11, 7, 8

(i) Construct the ANOVA table (10mks)

(ii) Using $\alpha = 0.05$, determine whether the treatments differ in their effectiveness.(6mks)

Question Three (20 marks)

a) Fit a second order polynomial to the following data (10mks)

i	1	2	3	4	5	6
x	0	0.5	1.0	1.5	2.0	2.5
y	0	0.25	1.0	2.25	4.0	6.25

b) Consider $y = \beta_0 + \beta_1 X + \varepsilon$, where y is financial aid and X is parental income. Both are measured in \$1,000's. A random sample 250 observations is drawn.

Obs	y_i	x_i
1	5.4	82.341
2	3.2	68.432
…	…	…
250	32	40.002
$\sum_{i=1}^{250}$	5,250	15,000

$$\sum x_i y_i = 270{,}900 \qquad \sum x_i^2 = 1{,}076{,}400 \qquad \sum y_i^2 = 405{,}370$$

i) Estimate the coefficients and interpret them. (3mks)

ii) The parents of a particular student have an income of \$92,000. Make an inference about the student's financial aid. (Answer in dollars and use $\alpha = 0.05$.) (4mks)

iii) Construct the 95% confidence interval for the slope. Interpret it. (3mks)

Question Four (20mks)

a) Researchers have sought to examine the effect of various types of music on agitation levels in patients who are in the early and middle stages of Alzheimer's disease. Patients were selected to participate in the study based on their stage of Alzheimer's disease. Three forms of music were tested: Easy listening, Mozart, and piano interludes. While listening to music, agitation levels were recorded for the patients with a high score indicating a higher level of agitation. Scores are recorded below.

Group	**Piano Interlude**	**Mozart**	**Easy Listening**
Early Stage Alzheimer's	21 24 22 18 20	9 12 10 5 9	29 26 30 24 26
Middle Stage Alzheimer's	22 20 25 18 20	14 18 11 9 13	15 18 20 13 19

Complete the following ANOVA table. (3mks)

Source	**SS**	**df**	**MS**	**F**
Type of Music	740	2		
Degree of Alzheimer's	30	1		
Music x Alzheimer's	260	2		
Within	178	24		
Total				

i. Graph your data. (2mks)

ii. Are there any significant main effects or an interaction effect. (3mks)

iii. Interpret your answer. (2mks)

b) In a comparison of the cleaning action of four detergents, 20 pieces of white cloth were first soiled with ink. The clothes were then washed under controlled conditions with 5 pieces washed by each of the detergents. Unfortunately three pieces of cloth were lost in

the course of the experiment. Whiteness readings made on the remaining 17 pieces of cloth are as shown.

Detergent			
A	B	C	D
77	74	73	76
81	66	78	85
61	58	57	77
76		69	64
69		63	

Assuming all whiteness readings to be normally distributed with common variance, test the hypothesis of no difference between the four brands as regards whiteness readings after washing.

(10mks)

Question Five (20mks)

c) A study examining differences in life satisfaction between young adult, middle adult, and older adult men and women was conducted. Each individual who participated in the study completed a life satisfaction questionnaire. A high score on the test indicates a higher level of life satisfaction. Test scores are recorded below.

Group	**Young Adult**	**Middle Adult**	**Older Adult**
Male	4 2 3 4 2	7 5 7 5 6	10 7 9 8 11
Female	7 4 3 6 5	8 10 7 7 8	10 9 12 11 13

Complete the following ANOVA table.

(3mks)

Source	**SS**	**df**	**MS**	**F**
Age	180	2		
Gender	30	1		
Age x Gender	0	2		
Within	44	24		
Total				

i. Graph your data.

(2mks)

ii. Are there any significant main effects or an interaction effect.

(3mks)

Interpret your answer.

d) Fit a second order polynomial to the following data (10mks)

i	1	2	3	4	5	6
x	0	0.5	1.0	1.5	2.0	2.5
y	0	0.25	1.0	2.25	4.0	6.25