JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY

SCHOOL OF HEALTH SCIENCES

UNIVERSITY EXAMINATION FOR DEGREE OF MASTER OF SCIENCE IN BIOSTATISTICS AND EPIDEMIOLOGY

1st YEAR 2nd SEMESTER 2018/2019 ACADEMIC YEAR

KISII CAMPUS

COURSE CODE: HES 5123

COURSE TITLE: ADVANCED BIOSTATISTICS

EXAM VENUE: **STREAM: (MSc. In BIOSTATISTICS & EPIDEMIOLOGY)**

DATE: **EXAM SESSION:**

TIME: 2.00 HOURS

Instructions:

1. **Answer any three 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 (20 MARKS)

a) The data below shows the haemoglobin levels (in g/decilitre) for patients with three sickle cell disease.

Sickle cell type	1	2	3	4	5	6	7	8	9	10	11
Hbss	7.2	7.7	8.1	8.3	8.5	8.6	8.7	9.1	9.1	9.8	10.3
Hbs/B-that	8.1	9.2	10	10.4	10.4	10.9	11.1	11.9	12	12.1	
Hbsc	10.7	11.3	11.5	11.6	11.7	11.8	12	12.3	13.3	138	

(i) What is the response variable and what is the factor variable (2 marks)

(ii) Obtain the fitted values and the residuals of

$\sum_i \sum_j e_{ij} = 0.$ (2 marks)

(iii) Obtain the ANOVA table for these data and use it to test at 5% whether there is significance difference in mean haemoglobin levels across the groups. (4 marks)

(b) Using an appropriate two sample t-test, evaluate whether haemoglobin levels for patients with Hbs and Hbsc are significantly difficult. (2 marks)

(c) suppose that we are interested in the factors that influence whether a political candidate wins an election. The outcome (response) variable is binary (1/0) for win /lose. The predictor variables of interest are the amount of money spent on the campaign (x_1), the amount of time spent campaigning negatively (x_2) and whether or not the candidate is incumbent (x_3). The regression coefficients are provided below;

x_1 = 0.8040
x_2 = -0.6754
x_3 = 0.0226
Null deviance = 499.98
Residual deviance = 358,52

i) Write the logistic regression formula and interpret the regression coefficients. (2 marks)

ii) Obtain the odds ratio for the regression coefficients showing clearly your working formula. (2 marks)

iii) Interpret the odds ratio. Also, compare performance of incumbent and non-incumbent using the odds ratio. (2 marks)

iv) Obtain pseudo-R squared and use it to evaluate the model goodness-of-fit interpreting accordingly. (2 marks)

d) A In a filariasis survey, the number of people with and without filariasis infestation in the two sex groups were as follows:

filariasis infestation	Male	Female	Total
Yes	28	20	48
No	237	222	459
Total	265	242	507

Test whether the prevalence of filariasis has statistical association with the sex using an appropriate test at 95% confidence level. (2 marks)

QUESTION TWO (20 Marks)

A company studied the effect of 3 different types of promotions on sales of its craters. Fifteen (n=15). Stores were selected at random for the study with 5 stores assigned to each promotional type. Data on the number of cases of product sold a previous promotional period (x) and on the current period (y) gives below.

Promotional type	1		2		3		4		5	
	Y_{i1}	X_{i1}	Y_{i2}	X_{i2}	Y_{i3}	X_{i3}	Y_{i4}	X_{i4}	Y_{i5}	X_{i5}
1	38	21	39	26	26	22	45	28	33	19
2	43	34	38	26	38	29	27	18	34	25
3	24	23	32	39	31	30	21	16	28	29

Fit an ANACOVA model to these data and assess whether there was promotional type effect on sales after adjusting for previous sales (use $\alpha = 0.05$)

QUESTION THREE (20 Marks)

a) Suppose a logistic regression model for the association between smoking and death is presented as follows:

$$log\left(\frac{p}{1-p}\right) = -7.5869 + 0.5522\ (CURSMOKE1) + 0.1181\ (AGE1) + 0.7759\ (MALE) + 0.6386\ (HIGHBP1) + 1.5834(DIABETES1)$$

i. Using the above model, what is the odds ratio of death for a 50 year old man who does not smoke, has high blood pressure and does not have diabetes (i.e. AGE1=50, CURSMOKE1=0, MALE=1, HIGHBP1=0, and DIABETES=0)? (3 marks)

ii. Does the answer to the previous question change if different values are set for AG1, MALE, HIGHBP1, and DIABETES? (3 marks)

iii. What is the model's estimate for the odds ratio of death for a diabetic (DIABETS1=1) compared to a non-diabetic (DIABETES1=0), controlling for MALE, HIGHBP1 and CURSMOKER1? (3 marks)

iv. What is the model's estimate for the odds ratio of death for a smoker (CURSMOKER1=1) compared to a non-smoker (CURSMOKER1=0), controlling for MALE, HIGHBP1 and DIABETES? (3 marks)

b) The following model contain the same risk factors listed in the previous model except that it does not include age

i. What is this model's estimate for the odds ratio of death for a smoker (CURSMOKER1=1) compared to a non-smoker (CURSMOKER1=0), controlling for MALE, HIGHBP1, and DIABETES1 (3 marks)

ii. Based on these two models, what conclusion can you reach about AGE1 being a confounder, when estimating the effect of smoking on the odds of dying, once you control for MALE, HIGHBP1, and DIABETES? (3 marks)

iii. Interpret the model intercepts when the other factors are held constant? (2 marks)

QUESTION FOUR (20 Marks)

The data below shows the number of cases of bread sold by a Bakery which wished to assess the effect of height of shelf display (factor A) and the width of shelf display (factor B) on the sales.

		Factor B (Display Width)	
		Regular	wide
Factor A (Height)	Bottom	47	46
		43	40
	Middle	62	67
		68	71

Top	41	42
	39	46

i) Obtain the fitted values (4 marks)

ii) Construct the ANOVA table (8 marks)

iii) Are there interactions effects (4 marks)

iv) Are there any main effects (4 marks)

(use $\alpha = 0.05$)

QUESTION FIVE (20 Marks)

The following tables show the code and sex-specific results from a prospective short study that examines the association between a binary exposure(E) and the development of a disease (D) during 20 years of follow-up.

a)

Full Data

	D +	D -	Total
E +	1123	8877	10000
E -	1008	8992	10000
Total	2131	17869	20000

Sex-specific data

Males

	D +	D -	Total
E +	259	1741	2000
E -	648	5352	6000
Total	907	7093	8000

Females

	D +	D -	Total
E +	864	7136	8000
E -	360	3640	4000
Total	1224	10776	12000

a) What is the value for the Crude Risk ratio, comparing exposed subjects to non-exposed subjects? (3 marks)
b) Using the Mentel-Haenszel formula, what is the value for the sex-adjusted Risk ratio, comparing exposed subjects to non-exposed subjects? (10marks)
c) Using the total data as standard population, what is the value for the standardized Risk ratio? (3 marks)
d) Using the risk ratio as a measure of association, is sex an effect modifies in this study? (4 marks)