JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY

SCHOOL OF HEALTH SCIENCES

UNIVERSITY EXAMINATION FOR DEGREE OF MASTER OF SCIENCE IN MSC PUBLIC HEALTH, EPIDEMIOLOGY AND DISEASE CONTROL

1st YEAR 2nd SEMESTER 2018/2019 ACADEMIC YEAR

KISII CAMPUS

COURSE CODE: HMP 5136

COURSE TITLE: STATISTICAL METHODS IN EPIDEMIOLOGY

EXAM VENUE: **STREAM: (MSc. In Biostatistics and Epidemiology; MSc. In Public Health, Epidemiology and Disease Control)**

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.**

QEUSTION ONE (20 MARKS)

a) In a random sample of 200 couples of bearing ages, it is found that 120 have adopted family planning method. In another random sample of 300 married couples of children bearing ages in rural areas, it is found that 150 have adopted family planning method. Do you think that the proportion of adopter couples in rural areas is significantly longer than proportion of adopter couples in the two samples together? (4 marks)

b) A randomly selected group of patients are investigated and classified according to their smoking habit and heart problems. The number of patients in each class is shown below.

Smoking habit	Heart problem		Total
	Yes	No	
Yes	55	15	70
No	10	25	35
Total	65	40	105

Do you think that smoking affects heart? (4 marks)

c) On a group of anaemic patients, an iron preparation was administered and haemoglobin levels of the patients before and after therapy were noted and are provided below. It is desired to find out whether there is a significant change in the haemoglobin level of the group after the therapy at $\alpha = 0.05$ level of significance. (6 marks)

Hb level in gm%

Patient number	Before therapy	After therapy
1	5.6	10.2
2	4.8	9.4
3	6.5	11.0
4	7.5	7.5
5	4.5	7.5
6	3.5	6.0
7	6.7	8.0
8	6.2	9.6
9	5.6	10.0
10	4.4	8.4
11	7.5	8.0
12	8.0	8.0

d) 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) 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. (6 marks)

QUESTION TWO (20 Marks)

The following is a summary of data presented by Refrigeration Company. The company had insisted to determine whether there was a relationship between the optimum Lot size of refrigeration replacement parts (x) and labour hours (y) required to produce the Lot.

Summary:

n=25; $\sum_{i=1}^{n} x_i$ = 1750; $\sum_{i=1}^{n} y_i$ = 7807; $\sum_{i=1}^{n} x_i\, y_i - n\bar{x}\,\bar{y}$ = 70670;

$\sum_{i=1}^{n}(x_i - \bar{x})^2$ = 19800; $\sum_{i=1}^{n}(y_i - \hat{y})^2$ = 54825

i) Obtain the regression parameter estimates (6 marks)

ii) Obtain a 95% confidence interval estimate for β_1 (4 marks)

iii) Test the hypothesis H_0: β_1 = 0 vs H_1: $\beta_1 \neq 0$ at 5% level of significance (3 marks)

iv) Obtain a 90% confidence interval estimate for the mean of the response when x_h =120 (3 marks)

v) Construct the ANOVA table (4 marks)

QUESTION THREE

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)

Screening for prostate cancer in men is a controversial topic. One of the most common screening mechanisms is the PSA test (prostate antigen test). In a meta-analysis, mistry and cable (2003) report that the sensitivity of the PSA test in 72.1% and the specificity is 93.2%. In Kenya, it is estimated that 16.1% of men will have prostate cancer of some part in their life (Kenya cancer society, 2012). Assume that prevalence of prostate cancer among men ages 75 and older in 16.1%. We examine the properties of the PSA screening test in men ages 75 and older, using the sensitivity and specificity value above.

a) Differentiate between sensitivity and specificity tests (2 marks)
b) Illustrate why screening test is necessary in clinical epidemiology giving details on a good screening program? (4 marks)
c) What is the probability of false negative test result? (5 marks)
d) What is the probability of false positive result? (5 marks)
e) What is the probability that a randomly selected man who is 75 years or older DOES NOT have prostate cancer, given that his PSA screening was positive? (4 marks)

QUESTION FIVE (20 Marks)

In a collective, the information on new born babies from two separate hospitals, the following information are recorded.

Hospital-1						
Gestation period x_{1j}	270	260	250	265	266	262
Birth-weight,$y_{1j}(in\ lb)$	7.0	6.5	6.2	7.0	6.1	5.8
Hospital-2						
Gestation period x_{2j}	263	265	270	255	260	260
Birth-weight,$y_{2j}(in\ lb)$	6.3	7.0	7.1	5.5	6.5	6.8

Test whether the influence of gestation period on both weight of babies in two hospitals are similar at $\propto=0.05$ levels of significance? (20 marks)