**JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY**

**SCHOOL OF HEALTH SCIENCES**

**UNIVERSITY EXAMINATION FOR DEGREE OF MASTER PUBLIC HEALTH**

**1ST YEAR 2ND SEMESTER 2018/2019 ACADEMIC YEAR**

# KISUMU LEARNING CENTRE

---

| | |
|---|---|
| **COURSE CODE:** | **HMP 5136** |
| **COURSE TITLE:** | **STATISTICAL METHODS IN EPIDEMILOGY** |
| **EXAM VENUE:** | **STREAM:** |
| **DATE:** | **EXAM SESSION:** |
| **TIME:** | **3.00 HOURS** |

---

**Instructions:**

1. **Answer any four Questions (Question One is Compulsary)**
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 1 (COMPULSORY)

a. When do you use Kappa statistics **(2 Marks)**

b. How can you assess if a variable is a confounder variable **(2 Marks)**
c. When do you use the Mantel-Haenszel test **(2 Marks)**
d. Name two common correlation measures **(2 Marks)**
e. Name the most convenient way of accessing the relationship between two numeric variables (**2 Marks**)

f. Assume you want to conduct a study on heart attack (the outcome) and blood levels of homocysteine. You wish to control possible confounding by smoking status (yes or no), blood pressure (as a continuous variable) and cholesterol level (as a continuous variable).

   i. Name the best statistical technique to accomplish this goal (**2 Marks**)

   ii. Give two main reasons for the choice of your test statistics (**3 Marks**)

## QUESTION 2

a. Outline the difference between sample and population in statistics giving two (2) merits and demerits of using any of them (**5 Marks**)
b. Define sampling error (**2 Marks**)
c. Descriptive statistics of serum HDL for neonates in Siaya and Kisumu counties in 2016 were as follows:

| Town | Mean serum HDL | Standard deviation | Sample |
|---|---|---|---|
| Siaya | 1.51 | 0.36 | 630 |
| Kisumu | 1.48 | 0.28 | 250 |

   i. List three (3) possible interpretations of the above results (**3 Marks**)
   ii. Determine whether the difference in serum HDL for neonates in Siaya and Kisumu counties is due to change or likely to be a real effect (**5 Marks**)

## QUESTION 3

a. The data below are from a case-control study of lung cancer, asbestos exposure and smoking status

| Smoking | Asbestos | Cases | Controls |
|---|---|---|---|
| Yes | Yes | 5 | 18 |
| | No | 10 | 72 |
| No | Yes | 75 | 20 |
| | No | 5 | 80 |

   i. Calculate the odds ratio of lung cancer associated with smoking and include a 95% confidence interval **(2.5 Marks)**

ii. Calculate the odds ratio of lung cancer associated with asbestos exposure and include a 95% confidence interval **(2.5 Marks)**

iii. Interpret results in i and ii **(2 Marks)**

iv. An investigator thinks it would be interesting to sort out the inter-relationship between asbestos, lung cancer and smoking by looking at the lung cancer risk associated with asbestos in smokers and non-smokers separately. Perform such a stratified analysis and report your findings **(4 Marks)**

v. Is confounding or interaction present **(2 Marks)**

## QUESTION 4

a. When do we use Mc-Nemar's test (**2 Marks**)

b. Sketch a graph showing no correlation between two variables but are dependent (**2 Marks)**

c. A study was carried out to examine variability in ultrasound examination for pathological lesions due to schistosoma mansoni infection. Two pathologists (A and B) independently examined the same 118 patients for pathology positive or negative. The results indicated that 36 patients were pathology positive by both pathologist A and B. 3 patients were pathology positive by pathologist B only, 16 patients were pathology positive by pathologist A only. The remaining 63 patients were pathology negative by both pathologist A and B.

   i. Construct a 2 x 2 contingency table (**3 Marks**)

   ii. State the alternative hypothesis (**2 Marks**)

   iii. Test the agreement or disagreement between the two pathologists (**6 Marks**)

Note: the **tabulated value of the test= 5.03 at 1 degree of freedom**

## QUESTION 5

a. List three (3) reasons for fitting a multiple regression instead of a simple regression model (**3 Marks**)

b. Logistic regression output below arises from a study of intensive care of adult patients in a hospital. The goal is to estimate a person's probability of death (Died=1, as opposed to survival Died=0), based on the person's age.

---

Variable | Coefficient. Standard Error z (coef.) P>|z| [95% Conf. Interval]

| | | | | | |
|---|---|---|---|---|---|
| Age | 0.028 | 0.011 | 2.61 | 0.009 | ? |
| Constant | -3.060 | 0.696 | -4.39 | 0.000 | ? |

i. Describe the relationship between age and probability of dying **(1 mark)**

ii. What is the equation that has been fitted to the data (**2 Marks**)

iii. Calculate odds ratio and it's 95 % confidence intervals for age and the constant (**5 Marks**)

iv. Is age a significant factor? Justify your answer (**2 Marks**)

v. What is the estimated probability of death for a patient who is 60 years old (**2 Marks**)

**QUESTION 6**

a. List four (4) requirements to determine sample size **(4 Marks)**

b. Assume that the sum of squares of blood pressure is 142, sum of squares of age is 140 and sum of products about the mean for blood pressure and age is 120.

i. Name the type of correlation measure to use to calculate the coefficient (**1 Mark**)

ii. Calculate the Pearson's correlation coefficient (**3 Marks**)

iii. Comment on the results obtained in (ii) above (**1 Mark**)

c. A group of surgeons wanted to conduct a trial on a new antibiotic whether it is effective in preventing gram-negative infections. Suppose they want to detect a reduction in the proportion that had infection (shock) from 10% to 5% or less and they are willing to tolerate a type 1 error of 5% and a 95% power.

i. What sample size should they use to achieve the above goal (**6 Marks**)