JARAMOGI OGIGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF HEALTH SCIENCES
MPH EXAMINATION
KISUMU CENTRE

END SEMESTER EXAMINATION
HMP 5136: STATISTICAL METHODS IN EPIDEMILOGY

December 2018

TIME ALLOWED: 2 hours 30 minutes.

Note: Answer question one and any other three questions

QUESTION 1 (COMPULSORY)

a. Define sampling error   **(2 Marks)**

b. What is R-squared important in regression analysis **(2 Marks)**

c. In a study of 88 births to women with a history of thrombocytopenia, the statistician attached to this study reported that the same condition was recorded in 20% of babies [95% CI: 13% to 30%].
    i. What interpretation would you give based on the above statement   (**3 Marks**)

d. A researcher wishes to find out if age, sex, occupation, body weight are associated with blood pressure.
    i.   What statistical test would you advise him to use? (**2 Marks**)

ii. Justify your answer given in (i) above (**4 Marks**)

e. When do we use McNamar's test during analysis (**2 Marks**)

QUESTION 2

a. Differentiate between Binomial and Poisson distribution (**3 Marks**)
b. Name two parameters used to describe binomial distribution (**2 Marks**)
c. The mean serum globulin level of a random sample of 400 healthy men in Kisumu County was found to be 30g/l with a standard deviation of 3g/l. If we assume serum globulin levels follow a normal distribution then,

    i. What proportion of healthy men in Kisumu county will have have serum globulin level below 27 g/l (**3 Marks**)

    ii. What proportion of healthy men in Kisumu county will have serum globulin level between 27g/l and 33g/l (**4 Marks**)

    iii. How many of healthy men in Kisumu county will have serum globulin level above 26 g/l (**3 Marks**)

QUESTION 3

a. List three (3) important assumptions of multiple linear regression (**3 Marks**)

b. Assume you want to test a difference in the mean responses for two treatments.
    i. Write down how you will set up the hypothesis (**2 Marks**)

c. A pediatrician examines the relationship between systolic pressure and age for a group of boys aged 5 to 13 years old. The data were analyzed with the following results

Source	SS	df	MS
Model	212.50	1	**?**
Residual	235.50	?	?
Total	?	8	

Number of obs = 9
$F_{(1, 7)}$ =?
Prob > F = 0.025
R-squared =?
Adj R-squared = 0.408

Variable	Coef.	Std. Err.	t	P>|t|	[95% Conf. Interval]
Age	2.08	0.82	2.54	0.04	?
Constant	41.92	7.70	6.75	0.0003	?

    i. Describe the relationship between blood pressure and age (**1 mark**)

    ii. Write down the equation that has been fitted to the data (**1 Mark**)

iii.   What is the blood pressure of a boy age 9 years (**2 Marks**)

iv.   Is the relationship statistically significant? Justify your answer (**2 Marks**)

v.   Complete the Analysis of Variance table (**4 Marks**)

QUESTION 4

a. Briefly describe two common correlation measures   (**4 Marks**)
b. A researcher wanted to assess the relationship between wing and tail length in centimeters among birds of a particular species as a result of that the following data were collected.

Wing length	Tail length
10.4	7.4
10.8	7.6
11.1	7.9
10.2	7.2
10.3	7.4
10.2	7.1
10.7	7.4
10.5	7.2
10.8	7.8
11.2	7.7
10.6	7.8
11.4	8.3

i.   Which method would you use to describe the relationship between wing and tail length   (**2 Marks**)

ii.   Assume that the association between wing and tail length is monotone, what type of correlation coefficient will you use to determine the strength of association between bird's wing and tail length   (**2 Marks**)

iii.   Calculate the correlation coefficient between wing and tail length   (**5 Marks**)

iv.   Interpret your answer obtained in (iii) above (**2 Marks**)

QUESTION 5

a. List four (4) requirements for determining sample size (**4 Marks**)
b. Briefly describe main tips of writing statistical analyses (**6 Marks**)

c. An investigator wanted to access whether health education can increase compliance rate with home management of malaria. Initial compliance is 35%. Suppose he want to detect an improvement of 25%. The investigator is willing to tolerate type I error of

2% and a 80% power. What sample size should he use to achieve the above goal (**5 Marks**)

QUESTION 6

a. Define sampling distribution   (**2 Marks**)
b. Name two (2) main reasons why sample size should be known before field work (**2 Marks**)
c. A public health official wishes to know how effective health education efforts are regarding smoking. Of 85 males sampled in 1900, 34 were found to be smokers. In 2012 a second random sample of 85 males similarly gathered, indicated that 25 were smokers
    i.   What is the percentage decrease in smoking during the period (**2 Marks**)
    ii.   Is there any significant change in smoking prevalence (**7 Marks**)
    iii.   Justify your answer in (ii) above (**2 Marks**)