*"Investing in Africa's Future"*

## FACULTY OF HEALTH SCIENCES

## SHS 205 INTRODUCTION TO EPIDEMIOLOGY AND BIOSTATISTICS

## END OF SEMESTER EXAMINATIONS

## APRIL/MAY 2019

## LECTURER: V. KAMPIRA

## DURATION: (3 HRS)

---

## *INSTRUCTIONS*

Answer **ALL** Questions in **Section A** and **ANY 3** questions from **Section B**

The mark allocation for each question is indicated at the end of the question

## SECTION A

**Question 1**

1a) Giving examples, distinguish between the following

(i)  Quantitative data and qualitative data.                                    [4]
(ii) Case control study and cohort study.                                       [4]

1b) Define and describe the three essential characteristics that are examined to study the cause(s) for disease in analytic epidemiology.                              [6]

1c) Define and describe the three essential characteristics of disease that we look for in descriptive epidemiology.                                                [6]

**Question 2**

a)  A case-control study of 1700 participants looked at the association between Tamoxifen and uterine cancer. The study included 689 cases.  There were 139 cases and 58 controls taking Tamoxifen.  Fill in the 2x2 table below.                        [3]

|           | Uterine Cancer | |
|-----------|-----|-----|
| Tamoxifen | Yes | No |
| Yes       |     |     |
| No        |     |     |

b) Calculate the odds of being a case among the exposed.                        [2]
c) Calculate the odds ratio of the above study.                                 [3]
d) Interpret the measure of association.                                        [2]

**Question 3**

a)  In a study of fingerprints, an important quantitative characteristic is the total ridge count for the 10 fingers of an individual. Suppose that the total ridge counts of individuals are approximately normally distributed with mean 140 and standard deviation of 50.
   (i)    Within what range of ridge count values will the middle 95% of the distribution lie?                                                                      [3]
   (ii)   What proportion of people will have a ridge count of at most130?        [2]
   (iii)  What proportion of people will have a ridge count of 190 or more?       [2]
   (iv)   Calculate the probability that a randomly selected individual will have a ridge count that is greater than the mean but smaller than 185.                [3]

## SECTION B

**Question 4**

a) Consider the following study:
In a study examining the relationship between oral contraceptives and bacteriuria, you follow women who do and do not use oral contraceptives over a three-year period, and find that 70 of the 500 individuals who use OC acquired bacteriuria, while 150 of 3000 individuals who don't use OC acquired bacteriuria.

(i) Is this a cohort or case-control study? [2]
(ii) What is the incidence among the exposed? [2]
(iii)What is the incidence among the unexposed? [2]
(iv)What is the relative risk? [2]
(v) What is the attributable risk? [2]
(vi)What is the attributable risk %? [2]

b) Among females in the united states between 18 and 74 years of age, diastolic blood pressure is normally distributed with mean $\mu=77$ mm Hg and standard deviation $\sigma=11.6$ mm Hg.
  (i) What is the probability that a randomly selected woman has a diastolic blood pressure less than 60 mm Hg? [2]
  (ii) What is the probability she has a diastolic blood pressure greater than 90 mm Hg? [2]
  (iii)What is the probability that the woman has a diastolic blood pressure between 60 and 90 mm Hg? [2]
  (iv)If a random sample of 500 females was collected from this population how many females would have a diastolic blood pressure that lies between 60 and 90 mm Hg? [2]

**Question 5**

a) You conduct a case-control study to examine the relationship between eating margarine and depression. You find that among the 185 patients who suffer from depression, 65 eat margarine, while 50 of the 230 controls eat margarine. Draw a 2x2 table and calculate the crude OR. [3]
b) Among the 100 women in this group who suffer from depression, 25 eat margarine. Among the 50 female controls, 5 eat margarine. Among the 85 male cases, 40 eat margarine. Among the 180 male controls, 45 eat margarine. Draw out the stratified 2x2 tables and calculate their respective ORs. [6]
c) Is this an example of effect modification or confounding? [2]
d) Explain in lay terms what this conclusion means. [2]
e) You want to calculate the average household expenditure on fast foods in monetary terms in an urban suburb infested with residents suffering from non communicable diseases. You sample 100 households and find that the mean in your sample is $300 with a standard deviation of $50.
  (i) What is the standard error of the sample? [2]

(ii) Calculate the 95% confidence interval around the estimate. [3]
(iii)Express in words the meaning of the 95% confidence interval. [2]

## Question 6

a) Describe and explain the following:
   (i) Type I error. [2]
   (ii) Type II error. [2]
   (iii)Power of a test. [2]
b) Outline the steps taken when a researcher needs to perform hypothesis testing using confidence intervals. [5]
c) What is the difference between confounding and effect modification? [5]
d) Define loss to follow up and briefly explain why is it important to minimise loss to follow up? [4]

## Question 7

a) Compare and contrast the likelihood of occurrence of selection and observation bias in case control studies and cohort studies. [6]
b) A study was conducted to investigate the relationship between maternal smoking during pregnancy and the presence of congenital malformations in the child. Among children with various types of defects other than oral cleft, 32.8% have mothers who smoked during pregnancy. In a random sample of 27 children born with oral cleft, 15 have mothers who smoked during pregnancy.
 (i) What is the point estimate p for this sample? [2]
(ii) Construct a 95% confidence interval for the population proportion and interpret it.[4]
(iii)You would like to know whether the proportion of mothers who smoked during pregnancy for children with oral cleft is the same as the proportion of who mothers who smoked for children with other defects. Conduct the test at 0.01 level of significance. [8]

# Question 8

a) What are the major options in the design of a study to minimize the occurrence of bias? [4]

b) State any form of bias and describe how it occurs. [2]

c) Below is some computer output:

**. regress weight height**

```
  Source |      SS           df       MS                     Number of obs =     92
---------+------------------------------             F( 1,   90) =  144.38
   Model | 31591.6473        1     31591.6473              Prob> F     = 0.0000
Residual | 19692.2222       90     218.802469              R-square    = 0.6160
---------+------------------------------             Adj R-square =  0.6117
   Total | 51283.8696       91     563.559006              Root MSE    = 14.792


-----------------------------------------------------------------------------
  weight |    Coef.        Std. Err.    t      P>|t|     [95% Conf. Interval]
---------+-------------------------------------------------------------------
  height | 5.091767       .4237488    12.016  0.000      4.249916   5.933618
   _cons | -204.7408      29.15972    -7.021  0.000      -262.6717  -146.8099
-----------------------------------------------------------------------------
```

(i) Identify the dependent and independent variables in this analysis. [2]

(ii) Write down the model fitted in this analysis. [2]

(iii) Interpret the coefficient of height from the model. [2]

(iv) Is there a statistically significant relationship between weight and height? [2]

(v) State and interpret the 95% confidence interval for the coefficient of height. [3]

(vi) Interpret the R-square value and differentiate it from the Adj R-square value. [3]