COLLEGE OF HEALTH, AGRICULTURE & NATURAL SCIENCES (CHANS)

DEPARTMENT OF HEALTH SCIENCES

SHS 203 INTRODUCTION TO EPIDEMIOLOGY AND BIOSTATISTICS

END OF SECOND SEMESTER EXAMINATIONS

APRIL/MAY 2019

LECTURER: MR V. KAMPIRA

DURATION: 3 HOURS

INSTRUCTIONS

Answer **all** questions in section A on separate answer sheets provided.
Answer any **3** questions in section B on separate answer sheets provided
Credit will be given for logical, systematic and neat presentations in sections B

SECTION A

Question 1

a) Define the following terms as used in bio-statistics and epidemiology
(i) Attack rate.   [2]
(ii) Epidemic.   [2]
(iii)   Endemic.
          [2]
(iv) Pandemic.   [2]
(v) P value.   [2]
b) Suppose a researcher wants to evaluate student behavior, knowledge and attitudes towards voluntary blood donations: A case of Africa University students.
(i) What sampling method must the researcher use, give reasons.   [3]
(ii) Which data collection tool is most appropriate and why?   [3]
(iii)   List any 4 variables that could be of importance.
          [4]

Question 2

1. Consider the following natality statistics for a given population in 2002. According to these data, the probabilities that a randomly selected woman who gave birth in 2002 was in each of the following age groups as given in Table 2.
   Table 2: Natality Statistics for a given population in 2002.

Age	Probability
<15	0.003
15 – 19	0.124
20 – 24	0.263
25 – 29	0.290
30 – 34	0.220
35 – 39	0.085
40 – 44	0.014
45 - 49	0.001
Total	**1.000**

   a) What is the probability that a woman who gave birth in 2002 was 24 years of age or younger?   [2]
   b) What is the probability that she was 40 or older?   [2]

c) Given that the mother of a particular child was under 30 years of age, what is the probability that she was not yet 20? [3]

d) Given that the mother was 35 years of age or older, what is the probability that she was under 40? [3]

Question 3

a) In a study, suppose you collected data that attempted to determine whether high cholesterol levels affect the frequency of heart disease. Two thousand eight hundred and fifty subjects with heart disease were compared to 2974 healthy subjects of similar age, and both groups were asked whether they were diagnosed with "high" cholesterol levels in the preceding 2 years. Among the 2850 heart disease patients, 358 were classified as having "high" cholesterol levels, while 229 of the 2974 disease-free subjects were recorded as having "high" cholesterol levels.

(i) Construct a 2x2 contingency table using the information given. [3]

(ii) Calculate the odds ratio. [3]

(iii)   Calculate the odds of being a case in the unexposed.
          [2]

(iv) Calculate the proportion of the diseased in the study. [2]

SECTION B

Question 4

a) A group of boys at Africa University in the faculty of health sciences indulge in drug abuse thought be attributable to peer pressure. Each of the boys either smokes or drinks alcohol or both. The probability that a boy smokes given that he drinks alcohol is 1/5. The probability that a boy drinks alcohol given that he smokes is 1/3. Let the event of smoking be S and the event of drinking alcohol be A.

(i) Find the probability that a boy selected at random is a smoker and drinks alcohol. [3]

(ii) Find the probability that a boy selected at random drinks alcohol only. [3]

(iii)   Find $P(A \cup S)$. [2]

b) In a certain hypothetical study of biological engineering, 2/3 of flower seeds have been treated to improve germination and 1/3 have been left untreated. The seeds which have been treated have a probability of germination of 0.8, whereas the untreated seeds have a probability of 0.5.

(i) Find the probability that a seed, selected at random, will germinate. [3]

The seeds were sown and given time to germinate

(ii) Find the probability that a seed selected at random had been treated, given that it had germinated. [3]

c) Creatinine reactive protein (CRP), a biomarker for heart disease was measured in men attending a hypertension clinic at Mutare provincial hospital. CRP is approximately normally distributed with mean equal to 2.4 mg/dl and a standard deviation equal to 0.8. A CRP level in excess of 3.0 mg/dl is considered abnormally high.

(i) What fraction of men will possess an abnormally high CRP level? [3]

(ii) What fraction of men will have a CRP level between 2.68 and 2.89? [3]

Question 5

a) To study rash illness among malaria spraying teams in Zimbabwe, Public health officers made a follow up of 126 individuals involved in the operation. Of the 30 subjects who developed rash, 25 had been exposed to the spraying chemical, whilst 31 of the exposed subjects had no rash.

(i) Construct a 2x2 table to indicate the study results. [3]

(ii) Calculate the appropriate measure of association. [4]

(iii) Compute a 95% confidence interval for the respective measure of association calculated in (ii). [4]

(iv) Is the measure of association statistically significant? Explain. [2]

b) In a cross-sectional survey administered to a random sample of 100 attendees of a local health fair, the following 2x2 table was constructed after reviewing the data:

	Diabetes		
Current Smoker	**Yes**	**No**	**Totals**
Yes	50	25	75
No	20	5	25
Totals	70	30	100

(i) Calculate the relative risk. [3]

(ii) Calculate the risk difference. [1]

(iii) Calculate the population attributable risk% and interpret. [3]

Question 6

a) Distinguish between the following

(i) Data and variable. [2]

(ii) Population and sample. [2]

(iii) Standard deviation and standard error. [2]

(iv) Type I and type II errors in hypothesis testing. [2]

(v) Discrete and continuous variable. [2]

b) What is an outbreak? [2]

c) Outline the steps and decisions taken in investigating an outbreak. [8]

Question 7

In a study to determine whether advice given by general practitioners (GPs) during a routine physical examination is effective in encouraging patients to stop smoking among current smokers, one group of patients was given a brief talk about the hazards of smoking and encouraged to quit. The second group received no advice pertaining to smoking. All patients were given a follow-up exam. In a sample of 114 patients who received the advice, 11 reported they had quit smoking. In the sample of 96 patients not given the advice, 4 had quit smoking.

a) Estimate the true difference in population proportions based on the differences between the sample proportions. [2]

b) Construct a 95% confidence interval for this difference. [4]

c) At the 0.05 level of significance, test the null hypothesis that the proportions of patients who quit smoking are identical for those who received advice and for those who did not. [8]

d) Do you believe that advice given by GPs is effective according to the results of this study? Why or why not? [2]

e) How does one control for confounding in the design stage of research. [4]

Question 8

a) List the assumptions of linear regression analysis. [4]

b) Write down the general regression equation and explain each term. [2]

c) Below is some computer out, a liner regression analysis was conducted.

```
. regress cholesterol time_tv

      Source |       SS       df       MS              Number of obs =     100
-------------+------------------------------           F(  1,    98) =   17.47
       Model |  5.04902329     1   5.04902329          Prob > F      =  0.0001
    Residual |  28.3220135    98   .289000137          R-squared     =  0.1513
-------------+------------------------------           Adj R-squared =  0.1426
       Total |  33.3710367    99   .337081179          Root MSE      =  .53759

-----------------------------------------------------------------------------
 cholesterol |     Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+---------------------------------------------------------------
     time_tv |   .0440691   .0105434     4.18   0.000     .0231461    .0649921
       _cons |  -2.134777   1.813099    -1.18   0.242    -5.732812    1.463259
-----------------------------------------------------------------------------
```

(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 time_tv from the model. [2]

5

(iv) Is there a statistically significant relationship between cholesterol and time_tv? [2]

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

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