



**COLLEGE OF HEALTH, AGRICULTURE & NATURAL  
SCIENCES**

**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**

1a Define the following terms as used in bio-statistics and epidemiology.

- i) Randomization. [2]
- ii) Standard error. [2]
- iii) Confounding. [2]
- iv) Risk difference. [2]
- v) Bias. [2]

1b) Describe any three data collection methods; give the advantages and disadvantages of each. [6]

1c) Distinguish between probability sampling and non probability sampling methods as used in research, Give examples. [4]

### **Question 2**

a) An investigator conducts a study to determine whether there is an association between caffeine intake and Parkinson's disease. He assembles 230 incident cases of PD and samples 455 controls from the general population. After interviewing all subjects, he finds that 64 of the cases had high daily intake of caffeine (exposed) prior to diagnosis and 277 of the controls had low daily intake of caffeine (unexposed) prior to the date of the matched case's diagnosis. Assemble the 2x2 table for this study using the information given. [3]

b) Calculate the odds of being a case among the exposed. [2]

c) Calculate the odds ratio for the above study. [3]

d) What does the odds ratio indicate? [2]

### **Question 3**

a) A random sample of 15 mothers at Sakubva antenatal clinic is taken from a population in which 40% had developed resistance to Nevirapine non-nucleoside reverse transcriptase inhibitor (NNRTI) used to treat HIV infection and AIDS and also used to prevent mother-to-child transmission of HIV in HIV-positive pregnant women. What is the probability that exactly 6 mothers in the sample are resistant to Nevirapine? [2]

b) The number of cases of tetanus reported in the United States during a single month in 1989 has a Poisson distribution with parameter  $\lambda = 4.5$ .

- i) What is the probability that exactly one case of tetanus will be reported during a given month? [2]
- ii) What is the probability that at most two cases of tetanus will be reported in a month? [2]
- iii) In a period of two months, what is the probability that four or more cases will be reported? [2]
- iv) What is the mean number of tetanus cases and the standard deviation of the cases reported in a one-month period? [2]

## **SECTION B**

### **Question 4**

- a) Name the bias that may arise from the situations described below:
- i) A case is more motivated to participate than a control, and thus more likely to report past exposures accurately. [2]
- ii) An interviewer learns to distinguish cases from controls, and subsequently differs slightly between them in how she asks her questions. [2]
- b) What preconditions are necessary for confounding? [4]
- c) Define and explain what is effect modification or (interaction). [3]
- d) Write brief notes on the following
  - i) Level of significance ( $\alpha$ ). [3]
  - ii) P value . [3]
  - iii) Intervention studies. [3]

### **Question 5**

You conduct a case-control study examining the relationship between drinking soda and colon cancer and find that among 1500 who have colon cancer, 400 drink soda, while among the 3000 controls who don't have colon cancer, 450 drink soda.

- a) Draw a 2x2 table and calculate the crude OR. [3]
- b) Now you stratify by gender and find the following:  
 Among women, 200 of 1000 who have colon cancer drink soda, while among the 2000 who don't have colon cancer, 300 drink soda.  
 Among men, 200 of the 500 who have colon cancer drink soda, but only 150 of the 1000 who don't have colon cancer drink soda.  
 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 simple terms what this conclusion means. [2]
- e) The distribution of systolic and diastolic blood pressure for female diabetics between the age of 30 and 40 have unknown means. However, their standard deviations are  $\sigma_s = 11.8 \text{ mm Hg}$  and  $\sigma_d = 9.1 \text{ mm Hg}$ , respectively.
  - i) A random sample of ten women is selected from this population. The mean systolic blood pressure for the sample is  $\bar{x} = 130 \text{ mm Hg}$ . Calculate a two sided 95% confidence interval for  $\mu_s$ , the true mean systolic blood pressure. [3]

- ii) Interpret this confidence interval. [1]
- iii) The mean diastolic blood pressure for the sample size of 10 is  $\bar{x} = 84$  mm Hg. Calculate a two sided 95% confidence interval for  $\mu_d$ , the true mean diastolic blood pressure. [3]

### Question 6

- a) List the assumptions of simple linear regression. [4]
- b) Below is some computer output:

. regress rifampicinresistance hivstatus						
Source	SS	df	MS	Number of obs = 311		
Model	.747058911	1	.747058911	F( 1, 309) = 9.04		
Residual	25.548761	309	.082682075	Prob > F = 0.0029		
Total	26.2958199	310	.084825226	R-squared = 0.0284		
				Adj R-squared = 0.0253		
				Root MSE = .28754		
rifampicin~e	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
hivstatus	.098782	.0328629	3.01	0.003	.0341186	.1634454
_cons	1.033851	.0256187	40.36	0.000	.9834422	1.08426

- i) Identify the independent and dependent variables in this analysis. [2]
- ii) Write down the model fitted in this analysis. [2]
- iii) Interpret the coefficient of HIV status in this model. [2]
- iv) Is there a statistically significant relationship between rifampicin resistance and HIV status? [3]
- v) State and interpret the 95% confidence interval of the coefficient of HIV status. [3]
- vi) Interpret the R-square value and differentiate it from the Adj R-square value. [4]

### Question 7

- a) The marks of 500 students in an examination are normally distributed with a mean of 45 marks and a standard deviation of 20 marks.
- i) Given that the pass mark is 41, estimate the number of candidates who passed the examination. [3]
- ii) If 5% of the students obtain a distinction by scoring  $\beta$  marks or more, estimate the value of  $\beta$ . [3]
- b) A researcher follows 200 women who exercise regularly and 300 women who do not exercise regularly. After 30 years of follow up, 20 of the women in the exercise group are diagnosed with osteoporosis while 30 women in the non exercise group are diagnosed with osteoporosis.

- i) Draw a 2x2 contingency table to illustrate the given information. [3]
- ii) Calculate the relative risk of developing osteoporosis between the two groups. [4]
- iii) Calculate the attributable risk%. [3]
- iv) Calculate the population attributable risk % and interpret. [4]

### Question 8

Suppose there are two treatments to lower cholesterol levels in subjects with elevated cholesterol, a standard drug and a new drug. You wish to determine whether the new drug is better than the standard drug. You randomly select from a population of subjects with elevated cholesterol and then randomly assign the subjects to either a standard drug or new drug. After six months on randomized therapy, the following data are collected.

	Standard drug	New drug
$\bar{x}$	<b>205</b>	<b>195</b>
$s^2$	<b>225</b>	<b>196</b>
$n$	<b>25</b>	<b>25</b>

- a) Using appropriate statistical notation, state the null and alternative hypothesis. [2]
- b) Calculate the value of the appropriate test statistic. [4]
- c) State the rejection region at the 5% level of significance. [2]
- d) Do you reject or do not reject the null hypothesis? Why or why not? [2]
- e) Find the p-value. [3]
- f) State your practical conclusion in light of the study. [2]
- g) Using information provided on the new drug, compute a 95% C.I for  $\mu$ , the mean cholesterol level after six months on the new drug therapy. [3]
- h) Interpret this confidence interval for your colleagues. [2]