



MERU UNIVERSITY OF SCIENCE AND TECHNOLOGY

P.O. Box 972-60200 – Meru-Kenya.

Tel: +254 (0)799529958, +254 (0)799529959, +254 (0)712524293

Website: www.must.ac.ke Email: info@must.ac.ke

University Examinations 2023/2024

SECOND YEAR, FIRST SEMESTER EXAMINATION THE DEGREE OF BACHELOR OF
SCIENCE IN STATISTICS

SMS 3215: INTRODUCTION TO EPIDEMIOLOGICAL RESEARCH METHODS

DATE: DECEMBER 2023

TIME: 2 HOURS

INSTRUCTIONS: *Answer question one and any other two questions*

QUESTION ONE (30 MARKS)

a) Define the following terms as used in epidemiology:

- i. Epidemiology (1 mark)
- ii. Odds (1 mark)
- iii. Specificity (1 mark)
- iv. Effect modification (1 mark)
- v. Risks (1 mark)

b) Given the 2 x 2 contingency table below:

	Unexposed	Exposed
Disease	9	17
No Disease	7	5

c) Compute:

- i. The fraction of cases with the disease among the exposed that is attributable to the exposure (2 marks)

- ii. The Odds ratio and its 95% confidence interval and interpret. (5 marks)
 - iii. Conduct a test to check whether the assumed exposure is associated with the disease outcome. (Take $\alpha = 0.05$) (6 marks)
- d) A study was conducted to evaluate the hypothesis that tea consumption and premenstrual syndrome are associated. A group of 188 nursing students and 64 tea factory workers were given questionnaires. The prevalence of premenstrual syndrome was 39% among the nursing students and 74% among the tea factory workers. How many people in each group have premenstrual syndrome? Arrange the data in a 2 by 2 table, hence obtain the risk difference and the proportion of risk attributable to the risk factor under study. (5 marks)
- e) A study had 34 patients in the control group, 4 were alive at the end of the study and of the 69 patients in the treatment group 24 were alive.

	Alive	Dead
Control	4	30
Treatment	24	45

Calculate:

- i) Proportion of all patients who died (1 mark)
- ii) The odds ratio and interpret your results (2 marks)
- iii) The relative risk and interpret your results (2 marks)
- iv) Discuss why the odds ratio is more commonly used as compared to the relative risk in epidemiology. (2 marks)

QUESTION TWO (20 MARKS)

- a) Distinguish between prevalence and incidence of a disease by their uses. (4 marks)
- b) Discuss at least 4 methods of controlling for confounding in epidemiological studies. (4 marks)
- c) In a town, 3,000 female residents had hypertension while 1,000 more developed the condition at the time of the study. Out of the 15,000 of the general female population in the town.
 - (i) Calculate the point prevalence and the incidence rates (2 marks)
 - (ii) Compute the 95% C.I for the prevalence and incidence rates (6 marks)

- d) An epidemiologist wants to study the efficacy of iron supplementation for the prevention of HIV infection. He wants to make sure that only subjects who are still free of HIV infection are enrolled in his trial. He therefore screens a large group of people using a diagnostic test. Based on the outcome of the diagnostic test, he decides who can participate in the trial.

Explain two vital measures that should be taken into consideration while picking out the subjects clearly distinguishing between the two and justifying the relevance of each in the study. (4 marks)

QUESTION THREE (20 MARKS)

- a) Distinguish between differential and non-differential misclassification process. (2 marks)
- b) Discuss any 3 types of measures of disease frequency. (3 marks)
- c) Ten percent of the new drugs manufactured by NOBSEH drug manufacturing company ltd turn out to be defective. Find the probability that a sample of 50 drugs chosen at random, exactly four will be defective by using Poisson distribution (4 marks)
- d) REMERA hospital authority has been keeping the records of the masses of patients for the past ten years. It has been shown that the masses are normally distributed with a mean of 72 kgs and Standard deviation of 10 kgs. What proportion of patients has masses;
- i. More than 82 kgs. (3 marks)
 - ii. Less than 65 kgs. (3 marks)
 - iii. Between 58 and 89 kgs. (5 marks)

QUESTION FOUR (20 MARKS)

- a) When are the following regression models used in epidemiological analyses and why are they preferred as opposed to the ordinary least squares model.
- i. Logistic regression. (2 marks)
 - ii. Poisson regression (2 marks)
- b) A study was carried out, where records of 210 patients who had undergone a procedure were reviewed. The data is summarized in the table below:

Had incident	wound	no wound
Appendectomy	infection	infection
Yes	7	124
No	1	78

Required:

- (i) Calculate the relative risk and 95% confidence interval for the relative risk (5 marks)
- (ii) Calculate the odds ratio and its 95% confidence interval and interpret the odds ratio (5 marks)
- (iii) Conduct a chi-square test for independence. Is there any evidence that the risk of wound infection is associated with having had an incidental appendectomy? (6 marks)
 $\alpha = 0.5, X_{1,0.05}^2 = 3.84$

QUESTION FIVE (20 MARKS)

- a) Discuss at least two types of epidemiological studies citing the advantages and disadvantages of each (4 marks)
- b) Consider the following two tests

		Culture test (Standard)		
		Positive	Negative	Total
Dipstick test	Positive	84	43	127
	Negative	10	92	102
Total		94	135	229

Compute and interpret:

- (i) Sensitive (1 mark)
- (ii) Specificity (1 mark)
- (iii) Positive predictive value (1 mark)
- (iv) Negative predictive value (1 mark)
- c) Birth defects were noted among women of maternal ages <20 to >24 and the following 2 x 2 table was developed

Birth defect	MATERNAL AGE		
	< 20 years	20 – 24 years	>24 years
Yes	101	105	36
No	1385	3755	6511

The table below is the output obtained after analyzing the above data using logistic regression and the odds ratio (OR) was also obtained

Output: Logistic Regression/odds ratio in R

Call:glm (formula = casegrp ~magecat 1+magecat2, family = binomial

Min	IQ	Median	3Q	max
-0.3752	-0.2349	-0.1050	-0.1050	3.2259

Coefficients:

	Estimate	std Error	z value	pr(> z)
Intercept	-5.1977	0.1671	-31.101	< 2e – 16 ***
magecat 1	2.5794	0.1964	13.137	< 2e – 16 ***
magecat 2	1.6208	0.1942	8.345	< 2e – 16 ***

Null deviance: 2364.1 on 11892 degrees of freedom

Residual deviance: 2148.6 on 11890 degrees of freedom

Aic: 2154.6

>Exp (c bind (OR coef(bd log2), confint (bdlog2))

	OR	2.5%	97.5%
Intercept	0.005529105	0.003910843	0.007541544
magecat 1	13.189149619	9.066531887	19.622868917
magecat 2	5.057367954	3.492376718	7.495720840

- (i) Write the equation of the fitted logistic regression model (2 marks)
- (ii) Calculate the predicted value (probability of birth defect) for women <20 year of age. (3 marks)
- (iii) Calculate the predicted value for women aged 20 – 24 (3 marks)
- (iv) Calculate the predicted value for women >24 years (2 marks)
- (v) Interpret the odds ratio output for this study (2 marks)