



KISII UNIVERSITY
UNIVERSITY EXAMINATIONS

SPECIAL/SUPPLEMENTARY EXAMINATIONS
SECOND YEAR EXAMINATION FOR THE AWARD OF THE
DEGREE OF BACHELOR OF SCIENCE IN
COMMUNITY NUTRITION AND DIETETICS/STATISTICS
SECOND SEMESTER, 2021/2022
(JULY, 2022)

STAT 205: BIOSTATISTICS

STREAM: Y2 S2

TIME: 2 HOURS

DAY: WEDNESDAY, 8:00 – 10:00 A.M.

DATE: 27/07/2022

INSTRUCTIONS

- 1. Do not write anything on this question paper.**
- 2. Answer Question ONE AND any other TWO Questions.**

SECTION ONE (30 MARKS)

a) Define the following terms as used in biostatistics citing an example in each.
(5marks)

- I. Treatment
- II. P-value
- III. Nominal measurement
- IV. Parameter
- V. Placebo

b) Which scale of measurement is most appropriate for the following variables:
(6marks)

- I. Political party affiliation (Democrat, Republican, unaffiliated).
- II. Highest degree obtained (none, high school, bachelor's, master's, doctorate).
- III. Patient condition (good, fair, serious, critical).
- IV. Hospital location (Nairobi, India, Kakamega, Kampala, Dar salam).
- V. Favorite beverage (beer, juice, milk, soft drink, wine, other).
- VI. How often do you feel depressed (never, occasionally, often, always).

c) According to recent UN figures, the annual gun homicide rate is 62.4 per one million residents in the United States and 1.3 per one million residents in the UK. (6marks)

i) Compare the proportion of residents killed annually by guns using the (i) difference of proportions, (ii) relative risk.

ii) When both proportions are very close to 0, as here, which measure is more useful for describing the strength of association? Why?

d) Each subject in a sample of 100 men and 100 women is asked to indicate which of the following factors (one or more) are responsible for increases in teenage pregnancy: A, the increasing gap in income between the rich and poor; B, the increase in the percentage of single-parent families; C, insufficient time spent by parents with their children. A cross classification of the responses by gender is shown below. (7marks)

Classification				
		A	B	C
Gender	Male	60	81	75
	Female	75	87	86

a) Is it valid to apply the chi-squared test of independence to this 2×3 table? Explain.

b) Explain how this table actually provides information needed to cross classify gender with each of three variables. Construct the contingency table relating gender to opinion about whether factor A is responsible for increases in teenage crime.

e) Based on diseases prevalence rates in Kenya, a survey has reported that the probability a newborn child eventually being a sick is 0.0263 for Urban males, 0.0049 for rural males, 0.0072 for rural females, and 0.0023 for urban females. (6marks)

i) Find the conditional odds ratios between region and whether a diseases prevalence, given gender. Interpret.

ii). If half the newborns are of each gender, for each region, find the marginal odds ratio between race and whether a murder victim.

QUESTION TWO (20 MARKS)

A doctor is investigating the effect of a woman's age on the success of an IVF (In Vitro fertilization) procedure. She has randomly selected 10 women aged under 35 and 10 women aged at least 35. From hospital records she has obtained the following data, which record the numbers of eggs obtained from the women and the numbers that were fertilized during one IVF procedure. She wants to investigate the effect of the woman's age on the probability of an egg being successfully fertilized. She calls this probability the "fertilization rate".

Women aged under 35		Women aged at least 35	
<i>Number of eggs</i>	<i>Number of Fertilized</i>	<i>Number of eggs</i>	<i>Number of fertilized</i>
10	9	7	6
9	7	10	7
7	5	9	5
5	3	8	4
10	9	6	4
7	7	5	1
9	5	7	4
8	8	6	4
7	2	5	2
7	5	7	5

a) Carry out a suitable exploratory analysis to see whether the fertilization rate might depend on the woman's age.

b) Let n_j denote the number of eggs and x_i the number of fertilized eggs for the i^{th} woman. Let t_j denote the fertilization rate for the i^{th} woman. Explain why a binomial distribution may be valid to model the data.

QUESTION THREE (20 MARKS)

a) For rate data, a GLM with identity link is

$$\mu/t = a + \beta x$$

Explain why you could fit this model using t and tx as explanatory variables and with no intercept or offset terms.

b.) Explain when we use the following models in biostatistics

i) Probit model

ii) Log linear models

iii) Logistic models

iv) General linear Models

QUESTION FOUR (20 MARKS)

A chi-squared variate with degrees of freedom equal to df has representation $Z_1^2 + \dots + Z_{df}^2$, where Z_1, \dots, Z_{df} are independent standard normal variates.

- a. If Z has a standard normal distribution, what distribution does Z^2 have?
- b. Show that, if Y_1 and Y_2 are independent chi-squared variates with degrees of freedom df_1 and df_2 , then $Y_1 + Y_2$ has a chi-squared distribution with $df = df_1 + df_2$.
- c. State three application of chi-square in biostatistics

QUESTION FIVE

Table below comes from one of the studies of the link between lung cancer and smoking. The study was done in 20 hospitals patients admitted with lung cancer in the previous year who were queried about their smoking behavior. For each patient admitted, researchers studied the smoking behavior of a non-cancer control patient at the same hospital of the same sex and within the same 5-year grouping on age.

A smoker was defined as a person who had smoked at least one cigarette a day for at least a year.

		Lung Cancer	
		cases	Control
Have smoked	Yes	688	650
	No	21	59
	Total	709	709

- a. Identify the response variable and the explanatory variable.
 - b. Identify the type of study this was.
 - c. Can you use these data to compare smokers with nonsmokers in terms of the proportion who suffered lung cancer? Why or why not?
 - d. Summarize the association, and explain how to interpret it.
- ii) Make notes on the following concepts:
- a) Incidental rate
 - b) Mortality rates
 - c) Prevalence rates