INSTRUCTIONS TO CANDIDATES

Answer Question One in and Any other TWO Questions

TIME: 2 Hours

QUESTION 1

(a) Describe a potential study of your choice that would have a categorical response variable. List variables that would be important in the study, distinguishing between response and explanatory variables. For each variable identify the measurement scale and indicate whether it would be treated as continuous or discrete. (20 marks)

(b) In the following examples, distinguish response and explanatory variables.
        i)   Heart disease(yes, No); Cholesterol level   (2 marks)
        ii)   Race (white, nonwhite); Gender (male, female); Vote for president
               (republican, democrat, other); Income   (3 marks)
        iii)   Hospital (A, B); Treatment (T_1, T_2); Patient outcome (survive, die)
                                        (2 marks)

   For each of the cases (i) to (iii) above , state the variable that would be treated as continuous (3 marks)

QUESTION 2

   The table below is taken from a research on the relationship between aspirin use and heart attacks by the physicians' Health Study Research Group of Kibabii university. The physicians Health study is a randomised clinical trial testing whether aspirin taken regularly reduces mortality from cardiovascular disease. Every other day, physicians participating in the study took either one aspirin or a placebo.

Myocardial Infarction

	Fatal attack	Non-fatal attack	No attack
Placebo	18	171	10,845
Aspirin	5	99	10,933

(a) Calculate the proportions with heart attacks for those taking,
        i)   Placebo   (2 marks)
        ii)   Aspirin   (2 marks)

(b) Obtain the difference in proportions and hence or otherwise, evaluate the relative risk. Give an interpretation of the relative risk so obtained in terms of heart attack and the treatment offered   (6 + 4 marks)

(c) Find the sample odds ratio   (6 marks)

QUESTION 3

Consider the following two-way table that categorizes a sample of people in a work force by income level (high or low) and education level (end after high school or end after college)

		Income Level	
		Low	High
Education Level	High School	a	b
	College	c	d

Use this table to express the following:

i) The estimated probability of low income for high school graduates (3 marks)

ii) The estimated probability of high income for high school graduates (3 marks)

iii) The odds favouring low income for high school graduates (3 marks)

iv) The odds favouring low income for college graduates (3 marks)

v) The odds ratio, $\hat{\theta}$ (4 marks)

Give an interpretation for the odds ratio, $\hat{\theta}$ (4 marks)

QUESTION 4

Consider the following data on the job status of students working within a university. The students are categorized as Freshmen or Sophomores and were required to confirm whether they had a job or not.

	Job Status	
	No	Yes
Freshmen	25	12
Sophomores	11	14

a) For Freshmen, obtain the odds favouring not having a job (3 marks)

b) For Sophomores, obtain the odds favouring not having a job (3 marks)

c) Calculate the odds ratio, $\hat{\theta}$ (5 marks)

d) Use the fact that $\dfrac{\ln(\hat{\theta}) - 0}{SE\left[\ln(\hat{\theta})\right]}$ has a distribution that is approximately standard normal for

large values of n to acertain whether there is an evidence of a strong association between job status and class category (9 marks)

QUESTION 5

(a) Distinguish Concordant from Discordant pairs giving a suitable example for clarity

(4 marks)

(b) Illustrate how proportions of Concordant and Discordant pairs could be used to measure the strength of the association between categorical variables   (4 marks)

(c) In a certain General Social Survey, the following data pertaining to job satisfaction and income was recorded

INCOME (Us $)	JOB SATISFACTION			
	Very dissatisfied	Little dissatisfied	Moderately satisfied	Very satisfied
< 6	20	24	80	82
6 - 15	22	38	104	125
15 – 25	13	28	81	113
>25	7	18	54	92

(d) Obtain,

      i)   The total number of concordant pairs, C   (3 marks)

      ii)   The total number of Discordant pairs, D   (3 marks)

From the values of C and D, comment on the status of income and job satisfaction

(2 marks)

(e) Calculate the value of gamma, $\hat{x}$ and use it to conclude about the nature of association between the categorical variables   (4 marks)