JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF HEALTH SCHENCES

UNIVERSITY EXAMINATION FOR THE MASTERS IN
BIOSTATICS & EPIDEMIOLOGY
1ˢᵗ YEAR   SEMESTER ONE 2019/2020 ACADEMIC YEAR
KISUMU

COURSE CODE: HMP 5114

COURSE TITLE: BIOSTATISTICS

EXAM VENUE: STREAM

DATE: 13/12/19   **EXAM SESSION: 2.00 – 5.00PM**

TIME: 3 HOURS

Instructions:

1. **Answer <u>ANY</u> 4 questions**

2. **Candidates are advised not to write on the question paper**

3. **Candidates must hand in their answer booklets to the invigilator while in the examination room**

Question 1

a. Which of these distributions is used for testing hypothesis? **(1/2 Marks)**

    i.   Normal distribution
    ii.   Chi-square distribution
    iii. Gamma distribution
    iv. Poisson distribution

b. A study was done to compare the lung capacity of coal miners to the lung capacity of farm workers. The researcher studied 200 workers of each type. Other factors that might affect lung capacity are smoking habits and exercise habits. The smoking habits of the two worker types are similar, but the coal miners generally exercise less than the farm workers (**2 Marks**)

    i. Identify the outcome variable of interest?
    ii. Is the outcome variable quantitative or qualitative?
    iii. What is the implied population?
    iv. What are the explanatory variables in this case?

c. 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 the 2X2 contingency table showing the disease on top and the exposure on the side. **(1 Marks)**
    ii. Calculate odds ratio (**OR**) & relative risk (**RR**) of developing osteoporosis between the two groups. (Show your work.) **(3 Marks)**

d. Categorize these measures according to the following level of measurements:
Nominal, ordinal, interval, or ratio (2.5 Marks)
    i. Time of first class
    ii. Major field of study
    iii. Course evaluation scale: poor, acceptable, good
    iv. Score on last exam (based on 100 possible points)
    v. Age of student

e. Hospital records show that of patients admitted suffering from Ebola, 80% die of it. What is the probability that of 6 randomly selected patient:-
    i. At least 4 will recover (**2Marks**)
    ii. Utmost 2 will recover (**1Marks**)

f. The time taken to assemble a car in a certain plant is a random variable having a normal distribution of 20 hours and a standard deviation of 2 hours. What is the probability that a car can be assembled at this plant in a period of time

a) less than 19.5 hours? **(2Marks)**

b) between 20 and 22 hours? **(2Marks)**

Question 2

a. The heights (in inches) of adult males in the JOOUST town campus are believed to be normally distributed with mean μ. The average height of a random sample of 25 students adult males is found to be $\bar{x} = 69.72$ inches, and the standard deviation of the 25 heights is found to be s = 4.15. Using a appropriate statistical probability distribution (*Hint*: n=25)

    i. Calculate 90% confidence interval for μ **(3Marks)**

    ii. Calculate 95% confidence interval for μ **(3Marks)**

b. A radar unit is used to measure speeds of cars on a motorway. The speeds are normally distributed with a mean of 90 km/hr and a standard deviation of 10 km/hr. What is the probability that a car picked at random is travelling at more than 100 km/hr? **(2Marks)**

c. Suppose that a rare disease has an incidence of 1 in 1000 person-years. Assuming that members of the population are affected independently, find the probability of k cases in a population of 10,000 (followed over 1 year) for k=0,1,2 **(6Marks)**

d. Define the following terms:-

    i. P-value **(1Mark)**

    ii. Level of significance **(1Mark)**

    iii. Level of confidence **(1Mark)**

Question 3

e. Suppose the researcher is interested in knowing protective effect of a drug on mortality in patients of myocardial infarction. He selected two groups of patients of myocardial infarction one group was given that drug and another group was given placebo. The both groups were kept under observation and at the end of study death in both groups were compared.

let us assume that previous study says that 20% of patient of myocardial infarction die within a specified time. The researcher feels that if the drug being tested increases survival to 30% then the finding can be considered as clinically significant. Effect size will be difference between proportions. $0.2 - 0.3 = -0.1$. At 5% of significance level and 80% power, the sample size will be? **(5Marks)**

f. In the French population, about 20% of people prefer Le Pen to other candidates (inc. Hollande and Sarkozy). An opinion poll asks 1000 people if they will vote for

Le Pen (*YES*) or not (*NO*). The expected number of Le Pen voters (*YES*s) in the poll is therefore $\mu = np = 200$

What is the standard deviation (approximately)? **(3Marks)**

g. Describe the statistical analysis you would use for the following situations and motivate your choice. For each of the situations, identify the **RESPONSE** variable and the **POPULATION** to be compared **(4.5 Marks)**

A) A travel company evaluates the group holidays it offers. To do this, the company asks the participants of the holidays an overall score between 1 and 10. The company offers holidays for groups of 15 persons to four areas in Europe (Norway, Italy, Ireland and Spain) and is interested in the differences in opinions between participants of these destinations.

B) A sport school offers a special activity program for people who want to lose weight. The weights of 25 participants of the program and 25 people who did not participate were measured at the beginning of the program, after 4 weeks and after 3 months to evaluate the effect of the program over time.

C) A school wants to measure the effect of extra lessons on the math results of male and female students. The school analyzes the exam results of 120 (60 male, 60 female) students who attended 0, 3 or 5 extra math lessons before taking the exam. 20 of each gender were randomly assigned to each of the extra lesson categories.

g. Name the bias that may arise from the situations described below:
- A case is more motivated to participate than a control, and thus more likely to report past exposures accurately
- An interviewer learns to distinguish cases from controls, and subsequently differs slightly between them in how she asks her questions

h. When is it **NOT** appropriate to use the odds ratio to approximate the relative risk? **(0.5Mark)**
   a. When the controls' exposure history is representative of *all* subject without disease
   b. When the case selection is representative of *all* subjects with disease (e.g. unbiased incident cases)
   c. When the study disease occurs frequently
   d. When the study outcome is rare

Question 4

a. A six-sided die is rolled 120 times with the following distribution of outcomes:

Outcome	Frequency
1	15
2	13
3	28
4	25
5	12

6	27

i. State the hypothesis and test the Chi-square goodness-of-fit **(2 Marks)**

b. In the table below, what is the fraction of cases with the disease among the exposed that is attributable to the exposure? **(3Mks)**

	Unexposed	**Exposed**
Disease	9	17
No disease	7	5

C. Which of the following is defined as the rule or formula to test a Null Hypothesis? **(1Mark)**
   A. Test statistic
   B. Population statistic
   C. Variance statistic
   D. Null statistic

b. The following is a classic 2x2 table:

	Case	Control
Exposed	a	b
Unexposed	c	d

a. What does a/b signify? (Hint: it's *an odds of...*) **(0.5Mark)**
b. What does c/d signify? **(0.5Mark)**
c. What is the formula, using this table, for calculating an odds ratio **(1Mark)**

d. Based on this table, and the formula from c, what would an odds ratio of 0.8 mean? 1.0? 3.2? **(3 Marks)**

e. A mixed box of 10 screws contains 5 that are galvanized and 5 that are non-galvanized. Three screws are picked at random without replacement. I want galvanized screws, so consider picking a galvanized screw to be a success.

Does the number of successes have a Binomial distribution? **(1 Mark)**

f. Do they follow **Poisson** distribution? Y/N, and give a reason **(4Marks)**

a. The number of heart attacks in Brighton each year

b. The number of planes landing at Hearthrow between 8 and 9am?

c. The number of cars getting a puncture at Thika road each year

d. The number of people in the UK flooded out of their homes in July

Question 5

a. A certain factory consumes on average 1000 m3 of water per day. A random sample of 100 days was taken to test if the mean daily water intake remains 1000 m^3 against the alternative that the mean water consumption has increased. We know that the sample mean equals $\_x = 1005$ m^3 and the sample quasi-variance is $S = 400$ m^3.

a) Define the null and alternative hypotheses for this test **(1Mark)**
b) Perform the test at the significance level $\alpha = 0:05$ **(2Marks)**
c) What is the smallest value of α for which the null hypothesis can be rejected? **(1Mark)**
d) Indicate which of the following statements are true/false and justify your answer:
   i. If we reject the hypothesis at the level $\alpha = 0:05$, we can also reject H0 at the level $\_\alpha = 0:1$ **(1Mark)**
   ii. The I Type error is the probability to reject the null hypothesis when H1 is true **(1Mark)**

   iii. If the p-value equals 0.15, we can reject the null hypothesis at the level 10%.**(1Mark)**

b. The TABLE below reports adjusted prevalence ratio of contextual and historical factors associated with age of sexual debut. Interpret the findings **(3Marks)**

Table 4. Age of first sex: adjusted prevalence ratios (N=292)[a], Kisumu contraceptive vaginal ring study, Kisumu, Kenya, 2014

Variable	Adjusted prevalence ratio <=15 vs. >15 yrs (95% confidence interval (CI))	p
Ever been inherited		
Yes	1.49 (1.16, 1.92)	0.002
No	Ref.	
Received gifts or favors as part of first sex		
Yes	1.38 (1.05, 1.82)	0.022

Variable	Adjusted prevalence ratio <=15 vs. >15 yrs (95% confidence interval (CI))	p
No	Ref.	
Type of contraceptive used in past 3 months		
DMPA	0.87 (0.61, 1.24)	0.437
COC	Ref.	

c. Decreasing the confidence level, while holding the sample size the same, will do what to the length of your confidence interval? **(1Marks)**

      i.   make it bigger

     ii.   make it smaller

   iii.   it will stay the same

   iv.   cannot be determined from the given information

d. If you increase the sample size and confidence level at the same time, what will happen to the length of your confidence interval? **(1Marks)**

  i.   make it bigger

 ii.   make it smaller

iii.   it will stay the same

iv.   cannot be determined from the given information

   e.   A 95% confidence interval for the mean number of televisions per American household is (1.15, 4.20). For each of the following statements about the above confidence interval, choose **true** or **false (3Marks)**

      i. The probability that μ is between 1.15 and 4.20 is .95.

      ii. We are 95% confident that the true mean number of televisions per American household is between 1.15 and 4.20.

     iii. 95% of all samples should have x-bars between 1.15 and 4.20. d) 95% of all American households have between 1.15 and 4.20 televisions.

Questions 6

    a. JOOUST School of Public Health believes that 28% of the applicants to that school have already given birth to at least 2 children. How large a sample is needed to

estimate the true proportion of students who have given birth in the department within 3% points with 99% confidence? **(6 Marks)**

b. A manufacturing process produces TV. tubes with an average life m=1200 hours and **s** = 300 hours. A new process is thought to give tubes a different life. And out of a sample of 100 tubes we find that they have an average life = 1265 hours. Is the new process any different from the old process? . Test the hypothesis using bth **p-value** & **critical value** method **(4Marks)**

c. The CEO of a large electric utility claims that 80 percent of his 1,000,000 customers are very satisfied with the service they receive. To test this claim, the local newspaper surveyed 100 customers, using simple random sampling. Among the sampled customers, 73 percent say they are very satisfied. Based on these findings, can we reject the CEO's hypothesis that 80% of the customers are very satisfied? Use a 0.05 level of significance **(4Marks)**

d. To investigate the effect of fruit on cholesterol, 40 participants were randomised into two treatmentgroups. Both groups were given the same information about healthy diet and activity, but one groupwas also instructed to eat 75 grams of dried apple every day. The cholesterol level for everyparticipant was measured (in mmol/L) before the study, and six months later, and the change wascalculated.

Choose the most appropriate procedure to decide if eating apple has any effect on the change incholesterol:
   i.   Chi-squared test for independence
   ii.   Paired T-test
   iii. Unpaired T-test
   iv.   Linear regression
   v.   McNemar's test.