



**JARAMOGI OGINGA ODONGA UNIVERSITY OF SCIENCE AND TECHNOLOGY
SCHOOL OF HEALTH SCIENCES**

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

COURSE CODE: HMP 5136

COURSE TITLE: STATISTICAL METHODS IN EPIDEMIOLOGY

EXAM VENUE: STREAM

DATE: 9.12.19

EXAM SESSION: 2.00 – 5.00PM

TIME: 3 HOURS

Instructions:

- 1. Answer ANY 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. Luos believe that eating fish makes them smarter, to test the hypothesis that eating fish makes one smarter, a random sample of 12 persons take a fish oil supplement for one year and then are given an IQ test. Here are the results:

116, 111, 101, 120, 99, 94, 106, 115, 107, 101, 110, 92

Test using the following hypotheses, report the test statistic with the P-value, then summarize your conclusion. **(4 Marks)**

$H_0: \mu = 100$

$H_a: \mu > 100$

- b. Which of these distributions is used for testing a testing hypothesis? **(1/2 Marks)**
- Normal distribution
 - Chi-square distribution
 - Gamma distribution
 - Poisson distribution
- c. 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)**
- Identify the outcome variable of interest?
 - Is the outcome variable quantitative or qualitative?
 - What is the implied population?
 - What are the explanatory variables in this case?
- d. 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.
- Draw the 2X2 contingency table showing the disease on top and the exposure on the side. **(1 Marks)**
 - Calculate odds ratio (**OR**) & relative risk (**RR**) of developing osteoporosis between the two groups. (Show your work.) **(3 Marks)**
- e. Categorize these measures according to the following level of measurements:
Nominal, ordinal, interval, or ratio (2.5 Marks)
- Time of first class
 - 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
- 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) 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 an appropriate statistical probability distribution (**Hint: $n=25$**)
- i. Calculate 90% confidence interval for μ **(3Marks)**
 - ii. Calculate 95% confidence interval for μ **(3Marks)**
- b. Suppose a study is performed concerning infant blood pressure and birthweight. All infants born in a specific hospital are ascertained within the first week of life while in the hospital and have their blood pressure measured in the newborn nursery. The researchers then divide the infants into two groups: a high-blood pressure group and a normal blood pressure group. They then compare birthweights between these two groups. What type of study design was used? **(1Mark)**
- i. Case-control study
 - ii. Cross-sectional study
 - iii. Randomized clinical trial
 - d. Cohort study
- c. A study looked at the effects of oral contraceptive use on heart disease in women 40 to 44 years of age. It found that among 5000 current oral contraceptive users at baseline, 13 women develop a myocardial infarction over a 3 year period, whereas among 10,000 non-oral contraceptive users, 7 develop a myocardial infarction over a 3-year period. What type of study design was used? **(1Mark)**
- i. Case-control study
 - ii. Cross-sectional study
 - iii. Randomized clinical trial
 - iv. Cohort study
- d. Consider the previous problem investigating the effects of oral contraceptive use on heart disease in women 40 to 44 years of age. Which of the following statistical tests is most appropriate for these data? **(1Marks)**
- i. binomial exact test
 - ii. chi-square test

- iii. McNemar's test
 - iv. independent samples t-test
- e. Researchers have conducted a survey of 1600 coffee drinkers asking how much coffee they drink in order to confirm previous studies. Previous studies have indicated that 72% of Americans drink coffee. The results of previous studies (left) and the survey (right) are below. At $\alpha = 0.05$, is there enough evidence to conclude that the distributions are the same?

Response	Frequency	% of coffee drinkers	
2 cups per wk	206	15%	
One cup per wk	193	13%	
One cup per day	462	27%	
2 + cups per day	739	45%	

State your hypothesis and calculate the Chi-square test statistics (Show your work) (6Marks)

Question 3

- a. Suppose a researcher wants to see the effect of a potential antihypertensive drug and He wants to compare the new drug with placebo. Researcher thinks that if this new drug reduces this blood pressure by 10 mmHg as compared to placebo then it should be considered as clinically significant. Let us assume standard deviation found in previously done studies was 25 mmHg. Suppose the researcher selects the level of significance at 5% and the power of study at 80%. and he thinks suitable statistical test in this condition will be two tailed unpaired t test. The effect size in this condition is 10 mmHg. Hence sample size will be (4Marks)
- 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 (YESs) in the poll is therefore
- $$\mu = np = 200$$

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

- g. A twin design is used to study age-related macular degeneration (AMD), a common eye disease of the elderly that results in substantial losses in vision. The study involves 66 sets of identical twins, where one twin has AMD and the other twin does not. The twins are given a dietary questionnaire to report their usual diet, and researchers' record whether each individual in the study takes multivitamin supplements. The data are shown in the following Table

	Normal Twin	
AMD twin	Multivitamin supplement	No multivitamin supplement
Multivitamin supplement	3	10
No multivitamin supplement	8	45

Which of the following statistical tests is most appropriate for these data? (3Marks)

- i. Fisher's exact test
- ii. chi-square test
- iii. McNemar's test
- iv. independent samples t-test

- g. When is it **NOT** appropriate to use the odds ratio to approximate the relative risk? (1Mark)

- 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. The side effects of a new drug are being tested against a placebo. A simple random sample of 565 patients yields the results below. At a significance level of $\alpha = 0.05$, is there enough evidence to conclude that the treatment is independent of the side effect of nausea? **(5Marks)**

Results	Drug	Placebo	Total
Nausea	36	13	49
No Nausea	254	262	516
Total	290	275	565

- a. 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

1. 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. In an investigation of pregnancy-induced hypertension, a group of 17 women with the disorder were treated with low-dose aspirin plus standard care. A second group of 24 women were given the standard care. In the aspirin group 13 women had a positive response to treatment. In the standard care only group 8 women had a positive response to treatment.

- i. Construct the 2 by 2 table used to test the null hypothesis that the fraction of positive responses is the same in each group. Label the rows and columns. You do not need to actually do the test **(2Mark)**
- ii. What type of test is used to test the null hypothesis in part (No need to explain your answer) **(1Mark)**
- iii. What assumptions are required for the hypothesis test in part **(2 marks)**

- a. 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)

Question 5

- b. A multiple-choice test question has four possible responses. The question is designed to be very difficult, with none of the four responses being obviously wrong, yet with only one correct answer. It first occurs on an exam taken by 400 students. The designers test whether more people answer the question correctly than would be expected just due to chance (i.e., if everyone randomly guessed the correct answer) (3Marks)

- i. Set up the hypotheses for the test.
- ii. Of the 400 students, 125 correctly answer the question. Find the P-value, and interpret.
- iii. Make a decision about H_0 , using $\alpha = .05$. Based on this decision, what can you conclude about the parameter?

- c. 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 Heathrow 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

- a. 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))
Ever been inherited	
Yes	1.49 (1.16, 1.92)
No	Ref.
Received gifts or favors as part of first sex	
Yes	1.38 (1.05, 1.82)
No	Ref.
Type of contraceptive used in past 3 months	
DMPA	0.87 (0.61, 1.24)
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

a. 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. The American Hospital Association reports that the mean cost to community hospitals per patient per day in U.S. hospitals was \$2035 last year. In that same year, a random sample of 25 Massachusetts hospitals yielded a mean daily cost of \$2100 with a sample standard deviation of \$160. It is known that the distribution of cost per patient per day is symmetric in the population of community hospitals. Is there evidence that that the cost of Massachusetts hospitals is different from the rest of the country? **(6 Marks)**
- State the null and alternative hypotheses (define any symbols you use) **(1Mark)**
 - Calculate the test statistic. **(2Marks)**
 - What is the distribution of the test statistic? ,no need to explain, just state **(1Mark)**
 - What is the degree of freedom? **(1Mark)**
 - Draw a picture of the p-value calculation **(1Mark)**
- b. A manufacturing process produces TV. tubes with an average life $m=1200$ hours and $s = 200$ 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 both **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. Researchers are interested in evaluating the effectiveness of a once a week educational program designed to prevent or stop smoking in teenagers. A simple random sample of 25 families with twin children in high-school is obtained. In each family, a randomly selected twin child is assigned to attend the educational program. The other twin does not attend. Two months after the end of the program the smoking status of each of the twins is assessed via a blood sample. The investigator would like to test whether the proportion of smokers in the twins who attended the educational program is different than the proportion of smokers in the twins who did not.
- Which statistical test will be most appropriate here? **(0.5Mark)**
 - State the hypothesis? **(0.5Mark)**