JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY

SCHOOL OF HEALTH SCIENCES

UNIVERSITY EXAMINATION FOR DEGREE OF MASTER PUBLIC HEALTH

1ST YEAR 1ST SEMESTER 2023/2024 ACADEMIC YEAR

KISUMU LEARNING CENTRE

COURSE CODE:	**HMP 5114**
COURSE TITLE:	**BIOSTATISTICS**
EXAM VENUE:	**STREAM:**
DATE:	**EXAM SESSION:**
TIME:	**3.00 HOURS**

Instructions:

1. **Answer any four Questions (Question One is Compulsary)**
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.**

SECTION A

Answer question one(Compulsary)

1. **Question one (10 marks).**
    a. Suppose that age in a population being considered has a normal distribution with a mean of 60 years and a standard deviation of 10. What proportion of the sample has age less than 50 years? (4 mark)
    b. Serum cholesterol was measured in a sample of 16 subjects before and after a six week diet. Data: mean difference 9.3 with standard deviation 16.8. Is the mean cholesterol lower than before? (3 mark)
    c. A sample of 18 people were drawn from a population of persons with Angina pectoris with a mean total cholesterol of 5.81mmol/l and standard deviation of 1.2, Calculate the standard error of the mean and approximate 95% confidence interval of the mean. (3 marks)

SECTION B

Answer any three Questions

2. Question five (20 marks).

To demonstrate t-test, the Dutch PCB/Dioxin study was set up to investigate adverse health effects of perinatal exposure to PBC's and dioxins. One of the topics that was studied was the influence of breast feeding on the development of the child. In this problem we look at the difference between breast-fed and formula-fed infants for two characteristics: age of the mother and LPCB, using a 2-sample t-test. The variable LCPB stands for the exposure to PCB's, qualified as the sum of four PCB congeners, ***log-transformed***. The SPSS output is given below.

	Group	N	Mean	Std. Deviation	Std. Error Mean
Age	breast-fed	88	28.99	4.314	.460
	formula-fed	95	28.88	3.179	.326
Lpcb	breast-fed	88	.3486	.16258	.01733
	formula-fed	95	.3996	.18772	.01926

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Age	Equal variances assumed	4.216	.041	.187	181	.852	.104	.557	-.995	1.204
	Equal variances not assumed	XXXX	XXXX	.185	159.237	.853	.104	.564	-1.009	1.218
Lpcb	Equal variances assumed	1.010	.316	????	????	????	????	.02605	????	????
	Equal variances not assumed	XXXX	XXXX	-1.970	180.203	.050	-.05104	.02591	-.10217	.00008

i. Which are the two assumptions underlying a 2 sample t-test? (2 marks)
ii. Comment on the validity of these assumptions for this application (2 marks)
iii. Is the difference in mean age statistically significantly different between breast-feeding and formula-feeding mothers?
    a. Give the p-value and formulate your conclusion (8 marks)
iv. Fill in the cells with a question mark in the result of the t-test for LPCB. (6 marks)
    a. What is your conclusion about the difference in LPCB between the groups? (2 marks)

3. Question two (20 marks).

Sixty-four (64) pregnant women at high risk of pregnancy-induced hypertension participated in a randomized controlled clinical trial comparing 100mg of aspirin daily and a matching placebo during the 3rd trimester of pregnancy. The observed numbers with hypertension are shown in the following table.

	Hypertension		
	Yes	no	Total
Aspirin	5	29	34
Placebo	10	20	30
Group Total	15	49	64

i. Is the risk of hypertension in aspirin treated women significantly lower than in placebo treated women?
    a. Which test do you use? Perform the test and give bound for the p-value? (4 marks)
    b. What is your conclusion about risk of hypertension in aspritin treated women? (1 marks)
ii. Give the **estimate** and **approximate** **95% confidence interval** for the following of hypertension between aspirin and placebo treated women
    a. Difference in risk (4 marks)
    b. Risk ratio (4 marks)
    c. Odds ratio (4 marks)
iii. Suppose a new study is planned. What sample size is approximately needed in order to have a power of 90% if the risk of hypertension is 0.1 lower in aspirin treated women ($\alpha = 0.05$)? (3 marks)

4. Question three (20 marks).

The data in the table below are form Charles Darwin's study of cross- and self-fertilization (Darwin C. (1876): The effect of cross-fertilization in the Vegetable Kingdom). Pairs of seedlings of the same age, one produced by cross-fertilization and the other by self-fertilization, were grown together so that the members of each pair were reared under nearly identical conditions. The data are the final height of each plant after a fixed period of time. The question is whether cross-fertilized reach higher final height than self-fertilized plants.

Cross-fertilization	Self-fertilization	Difference
23.5	17.4	6.1
12	20.4	-8.4
21	20	1
22	20	2
19.1	18.4	0.7
21.5	18.6	2.9
22.1	18.6	3.5
20.4	15.3	5.1
18.3	16.5	1.8
21.6	18	3.6
23.3	16.3	7
21	18	3
22.1	12.8	9.3
23	15.5	7.5
12	18	-6

Descriptive Statistics

	N	Mean	Std. Deviation
Cross-fertilized	15	20.1933	3.61613
Self-fertilized	15	17.5867	2.03816
Difference	15	2.6067	4.71282
Valid N (listwise)	15		

i. Which 2 non-parametric tests are appropriate for analysis of these data? (2 marks)
   a. Formulate the null hypothesis (2 marks)

ii. Test the null hypothesis that cross-fertilized and self-fertilized plans reach on average the same final height. (6 marks)
   a. Give also a 95% confidence interval for the mean difference in final height. (6 marks)
   b. What is the underlying assumption for this test and confidence interval? (4 marks)

5. Question four (20 marks).

In an experimental study (Smith, 1990) on the effect on recurrent myocardial infarction of long term anticoagulant therapy after myocardial infarction, 607 patients with anticoagulant therapy were compared with a non-coagulated group of the same size. The numbers of patients experiencing a recurrence in these groups were, respectively, 82, and 124 during follow-up after myocardial infarction. Answer the following questions.

   d. Does anticoagulant therapy lead to a significant reduction in recurrent myocardial infarction?
      i. Calculate the appropriate test statistic and give limits for the p-value. (4 mark)

e. Calculate an estimate for the relative risk (=risk ratio) (RR) of experiencing a recurrent myocardial infarction in anti-coagulated patient relatively to non-anti-coagulated patients. (4 mark)
f. Calculate an approximate 95% confidence interval for RR and explain the (in)-consistency with your answer. (4 mark)

Another similar study (Ebert 1969) was carried out in two treatment groups: 385 anti-coagulated patients and 350 non-anti-coagulated patients. The respective numbers of re-infarctions were 60 and 72. Hence, in this study an estimate for ln(RR) is -0.278 with an estimated variance of 0.0251.

g. Test the null hypothesis that RR is the same in both studies and calculate the estimate for RR using the results of both studies (3 mark)
h. What is the pooled RR and its 95% confidence interval, Does anticoagulant therapy lead still leads to a significant reduction in recurrent myocardial infarction when the data is pooled? (5 mark)