CHUKA  **UNIVERSITY**

UNIVERSITY EXAMINATIONS

EXAMINATION FOR THE AWARD OF DEGREE OF BACHELOR OF SOCIOLOGY

SOCI 475: SOCIAL STATISTICS II

STREAMS: SOCI, CB4, CB5, CB7, CB24 **TIME: 2 HOURS**

DAY/DATE: WEDNESDAY 20/12/2023 **11.30 A.M. – 1.30 P.M.**

INSTRUCTIONS

i. Answer question ONE and any other two questions
ii. Use illustrations where appropriate

1. a) Define the following concepts

   i) Parametric test (2 marks)

   ii) Confidence level (2 marks)

   b) Differentiate between:

   i) correlation and regression (4 marks)

   ii) independent samples t-test and one-way ANOVA (4 marks)

   c) Illustrate the types of relationships depicted by a correlation coefficient. (6 marks)

   d) Discuss the four possible outcomes of hypothesis testing? (4 marks)

   e) A researcher wants to know if there is a difference in how busy someone is based on whether that person identifies as a morning person or a night person. The researcher gathers data from people in each group, coding so that higher scores represent higher levels of being busy. Test the investigator's theory using an alapha level of 0.05.

Sample	Mean	SD	N
Night person	19.50	6.14	8
Morning person	26.67	3.39	9

i) State both the null and alternative hypothesis (2 marks)

ii) Find the critical value of the appropriate statistic. (2 marks)

iii) Compute the statistic (4 marks)

2. a) A study is run to evaluate the effectiveness of an exercise program in reducing blood pressure (BP) in patients with hypertension. A total of 15 patients with hypertension enroll in the study, and their BPs are measured. Each patient then participates in an exercises training program where they learn proper techniques and execution of a series of exercises. Patients are instructed to do the exercise program 3 times per week for 6 weeks. After 6 weeks, BPs are again measured. The data are shown below and are not normally distributed.

Patient	BP before exercise program	BP after exercise program
1	125	118
2	132	134
3	138	130
4	120	124
5	125	105
6	127	130
7	136	130
8	139	132
9	131	123
10	132	128
11	135	126
12	136	140
13	128	135
14	127	126
15	130	132

Compute a Wilcoxon Signed Rank Test to determine whether there is a difference in BP after exercise program as compared to before? (10 marks)

b) A school administrator wishes to use celebrity influence to encourage students to make healthier choices in the school cafeteria. Every day at lunch time students get their lunch and a drink in three separate lines leading to three separate serving stations. As an experiment, the school administrator displayed a poster of a celebrity drinking milk at each of the three areas

where drinks are provided, except the milk in the poster is different at each location: white milk, strawberry-flavoured pink milk, and chocolate milk. After the first of the experiment the administrator noted the students' milk choices separately for the three lines. The data are given in the table provided.

Poster choice	Student lines		
	Line 1	Line 2	Line 2
White milk	38	28	40
Strawberry-flavored pink milk	18	51	24
Chocolate milk	32	32	53

Test, at the 1% level of significance, whether there is sufficient evidence in the data to conclude that the posters had some impact on the student's drink choices. (10 marks)

3. a) The following data related the rubber percentage of two types of rubber plants, where the samples have been drawn independently.

Type 1	6.21	5.7	6.04	4.47	5.22	4.45	4.84	5.84	5.88	5.82	6.09	5.59
	6.06	5.59	6.74	5.55								

Type 2	4.28	7.71	6.48	7.71	7.37	7.20	7.06	6.4	8.93	5.91	5.51	6.36

Test for their mean difference at 5% level of significance? (10 marks)

b) A chicken farmer tested three kinds of poultry feeds with the weights (kgs) of the grown chicken in each sample given below.

Feed A	Feed B	Feed C
12.3	12.1	11.5
11.4	13.4	12.4
13.4	14	10.4
12.5	13.6	12.6
12	12.8	11.8
13.1	14.2	11.9

Test at $\alpha = 0.05$ whether there is a difference in the mean weights of chicken consuming the different feeds. Assume the distributions are normal and variances are equal. (10 marks)

4. a) A botanist prepares 16 identical planting pots and then put different types of soils into the pots. A tomato seedling is transplanted into each plot. Here are data on the increase in height of the seedlings (in centimeters) 16 days after planting.

Soil type	Seedling growth			
Loam	10.8	9.1	13.5	9.2
Black cotton	11.1	11.1	11.3	8.2
Clay	5.4	4.6	7.4	5.0
Alluvial silt	5.8	5.3	3.2	7.5

i) State the hypothesis for the test. (2 marks)

ii) Does soil type affect growth? (apply the Kruskal – Wallis test) (6 marks)

iii) What do you conclude? (2 marks)

b) Using the income data from 1991 GSS, calculate a t test statistic to determine if the difference between the two group means is statistically significant. (6 marks)

	Mean	S.D	N
Men	22,0252.51	17,734.92	434
Women	14,331.21	12,165.89	448

c) Differentiate between paired sample and independent sample t-test (4 marks)

5. The annual number of driver deaths per 100,000 data for the midpoints of different age groups is as summarized below.

Age	Number of driver deaths per 100,000
17.5	38
22	36
29.5	24
44.5	20
64.5	18
80	28

Using “ages” as the independent variable and “Number of driver deaths per 100,000” as the dependent variable.

i) Draw a scatter plot to show the distribution (4 marks)

ii)	Calculate the least squares regression line	(6 marks)
iii)	Predict the number of deaths for ages 40 and 60	(2 marks)
iv)	Find the correlation coefficient	(4 marks)
v)	Is the correlation coefficient significant?	(4 marks)

…….