CHUKA   **UNIVERSITY**

UNIVERSITY EXAMINATIONS

EXAMINATION FOR THE AWARD OF DEGREE OF BACHELOR OF ARTS

GEOG 246: QUANTITATIVE METHODS IN GEOGRAPHY

STREAMS: BA Y2S2   **TIME: 2 HOURS**

DAY/DATE: MONDAY 12/07/2021   **2.30 P.M. – 4.30 P.M.**

INSTRUCTIONS:
  • **Answer question ONE and any other TWO questions.**

1.   (a)   Define the following:
             (i)   Central Limit Theorem   (2 marks)
             (ii)   Standard error of the mean   (2 marks)
             (iii)   Degrees of freedom   (2 marks)

      (b)   Differentiate between:

             (i)   Parametric test and non parametric test   (4 marks)

             (ii)   Critical value and critical region   (4 marks)

      (c)   Suppose your computed r=-0.71 with 17 data points, test the significance of the
             correlation coefficient.   (4
marks)

      (d)   Which statistical analysis is appropriate to analyze the following: "Comparing the
             average number of hours per week spent on Facebook during the first
week in   April and the first week in May for randomly selected students at
Chuka   University, measured on the same 100 students"
      (2 marks)

      (e)   What assumptions should variables meet for the Kruskal Wallis test. (4 marks)

      (f)   An investigator theorizes that people who participate in a regular program of
             exercise will have levels of systolic blood pressure that are significantly

different from that of people who do not participate in a regular program of exercise. To test this idea the investigator randomly assigns 21 subjects to an exercise program for 10 weeks and 21 subjects to a non-exercise comparison group. After ten weeks the mean systolic blood pressure of subjects in the exercise group. After ten weeks the mean systolic blood pressure of subjects in the exercise group is 137 and the standard deviation of blood pressure values in the exercise group is 10. After ten weeks, the mean systolic blood pressure of subjects in the non-exercise group is 9.0. Test the investigator's theory using an alpha level of 0.5.

(i) State both the null and alternative hypotheses. (2 marks)
(ii) Find the critical value of the appropriate statistics. (2 marks)
(iii) Compute the statistic. (4 marks)
(iv) Make a decision about the null hypothesis and state the conclusion. (2 marks)

2. (a) A new chemotherapy treatment is proposed for patients with breast cancer. Investigators are concerned with patient's ability to tolerant the treatment and chemotherapy assess their quality of life both before and after receiving the new treatment. Quality of life (QOL) is measured on an ordinal scale and for analysis purposes, numbers are assigned to each response category as follows: 1=Poor, 2=Fair, 3=Good, 4= Very Good, 5=Excellent. The data are shown below.

Patient	QOL Before Chemotherapy Treatment	QOL After Chemotherapy Treatment
1	3	2
2	2	3
3	3	4
4	2	4
5	1	1
6	3	4
7	2	4
8	3	3
9	2	1
10	1	3
11	3	4
12	2	3

Is there is a difference in QOL after chemotherapy treatment as compared to before? (10 marks)

(b) A random sample of 400 people were surveyed and each person was asked to report the highest education level they obtained. The data that resulted from the survey is summarized in the following table:

High School   Bachelors   Masters   Ph.D.   Total

Female	60	56	46	42	204
Male	40	46	53	57	196
Total	100	102	99	99	400

Are gender and education level dependent at 5% level of significance? In other words, given the data collected above, is there a relationship between the gender of an individual and the level of education that they have obtained? (10 marks)

3. (a) A clinical trial is run to assess the effectiveness of a new anti-retroviral therapy for patients with HIV. Patients are randomized to receive a standard anti-retroviral therapy (usual care) or the new anti-retroviral therapy and are monitored for 3 months. The primary outcome is viral load which represents the number of copies per milliliter of blood. A total of 30 participants are randomized and data are shown below.

Standard Therapy	New Therapy
7500	400
8000	250
2000	800
550	1400
1250	8000
1000	7400
2250	1020
68000	6000
3400	920
6300	1420
9100	2700
970	4200
1040	5200
670	4100
400	Undetectable

Is there statistical evidence of a difference in viral load in patients receiving the standard versus the new anti-retroviral therapy? (10 marks)

(b) Four brands of flashlight batteries are to be compared by testing each brand in five flashlights. Twenty flashlights are randomly selected and divided randomly into four groups of five flashlights each. Then each group of flashlights uses a different brand of battery. The lifetimes of the batteries, to the nearest hour, are as follows.

Brand A	Brand B	Brand C	Brand C
42	28	24	20
30	36	36	32
39	31	28	38

| 28 | 32 | 28 | 28 |
| 29 | 27 | 33 | 25 |

Preliminary data analyses indicate that the independent samples come from normal populations with equal standard deviations. At the 5% significance level, does there appear to be a difference in mean lifetime among the four brands of batteries?   (10 marks)

4.   (a)   A botanist prepares 16 identical planting pots and then introduces different numbers   of nematodes (microscopic worms) 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.

Nematodes	Seedling growth			
0	10.8	9.1	13.5	9.2
1,000	11.1	11.1	11.3	8.2
5,000	5.4	4.6	7.4	5.0
10,000	5.8	5.3	3.2	7.5

(i)   State the hypotheses for the test.   (2 marks)
(ii)   Do nematodes appear to retard growth? (Apply the Kruskal-Wallis test) (6 marks)
(iii)   What do you conclude?   (2 marks)

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

	Mean	S.D	N
Men	22,052.51	17,734.92	434
Women	14,331.21	12,165.89	448

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

5.   (a)   Suppose you computed r=-0.624 with 14 data points, test the significance of the correlation coefficient.   (6 marks)

(b)   Deep sea drivers have maximum dive times that they cannot exceed when going to   different depths. The data in the table show different depths with the maximum dive   times in minutes.

X (depth in feet)	Y (maximum dive time)
50	80
60	55
70	45
80	35
90	25
100	22

(i)   Draw a scatter plot to show the distribution.   (5 marks)

(ii)   Calculate the least squares regression line.   (7 marks)

(iii)   Predict the maximum dive time for 110 feet.   (2 marks)