(Knowledge for Development)

KIBABII UNIVERSITY

UNIVERSITY EXAMINATIONS

2021/2022 ACADEMIC YEAR

FOURTH YEAR FIRST SEMESTER

MAIN EXAMINATION

FOR THE DEGREE OF BACHELOR OF SCIENCE

COURSE CODE: STA 415

COURSE TITLE: NON PARAMETRIC METHODS

DATE: 18/05/2022 **TIME:** 2:00 PM – 4:00 PM

INSTRUCTIONS TO CANDIDATES

Answer Question One and Any other TWO Questions

TIME: 2 Hours

This Paper Consists of 5 Printed Pages. Please Turn Over.

QUESTION 2:

a) The following are the scores of certain randomly selected students at mid-term (MT) and final examinations.

MT scores X	55	57	72	90	57	74	
Final score Y	80	76	63	58	56	37	75

The hypothesis H_0 that the distribution of scores at two occasions is the same against H_1

i.e. $H_0: F_Y(x) = F_X(x)$ Vs $H_1: F_Y(x) \neq F_X(x)$

Use the Mann-Whitney U-test. (10 marks)

b) . On a lonely country road, the number of vehicles passing a particular spot is noted for 60 consecutive minute as follows.

Number of vehicles	0	1	2	3	4	5
Number of minutes	25	15	10	5	3	2

Test $H_0: F(x) = F_0(x)$

Where $F_0(x)$ is a Poisson distribution at 5% level of significance.

Hint: λ can be estimated using the above:

λ = mean number of vehicles per minute.

$$\lambda = \frac{Number\ of\ vehicles}{number\ of\ minutes} = \frac{15}{60} = 0.25$$ (10 marks)

QUESTION 3:

(a)The following data, in tones are the amounts of sulfur oxides emitted by a large industrial plant in 40 days:

17 15 20 29 19 18 22 25 27 9

24 20 17 6 24 14 15 23 24 26

19 23 28 19 16 22 24 17 20 13

19 10 23 18 31 13 20 17 24 14

QUESTION 1:

(a) (a) Non parametric tests are also called distribution free tests, explain why it is so. (1 mark)

(b) State the non-parametric alternative to :

(i) One sample t- test (2 marks)

(ii) Two –sample t-test (2 marks)

(iii) One-way analysis of variance(ANOVA) (2 marks)

(iv) Two-way analysis of variance (ANOVA) (2 marks)

(c)Let $x_1, x_2, x_3, \ldots\ldots x_n$ be a random sample of size n from a population of cumulative distribution function, F(x) .Let x_r be the r-th order statistic. Define $F_r(X) = P(x_r \leq x)$ to be the cumulative distribution of the r-th order statistic. Obtain an expression of the probability density function $f_r(x)$ and cumulative distribution function, $F_r(x)$ for this order statistic for a case when,

i) r=1

ii)r=n

(d) Suppose that the light of a certain bulb is exponentially distributed with mean 100 hours.

$$f(x) = \frac{1}{100} e^{-\frac{1}{100}x} \qquad x \geq 0$$

If ten such light bulbs are installed simultaneously, what is the distribution of the life of the bulb that fails first and what is its expected life.

(e) Let $x_1, x_2, \ldots, x_n$ be independent observations from the uniform distribution in the interval (0,1). Obtain $f_1(x)$, $f_n(x)$ and $f_r(x)$

Hint: For a uniformly distributed random variable in the interval (a,b), $f(x) = \frac{1}{b-a}$

(f) Let $x_{(r)}$ be r^{th} order statistic, then

$$f_r(x) = \frac{n!}{(r-1)!(n-r)!}[F(x)]^{r-1}[1 - F(x)]^{n-r} f(x)$$

Take *x to* be a random sample from a uniform distribution over (0, 1), obtain

i) $E(x_{(r)}^k)$ for $k \geq 1$.

ii) $E(x_{(r)})$

iii) $E(x^2_{(r)})$

Use the sign test to test,

H_o: $\mu=21.5$ against H_1: $\mu< 21.5$ at the 0.05 level of significance. (10 marks)

(b)The following are the final examination grades of samples from three groups of students who were taught statistics by three different methods (classroom instruction and laboratory work, only classroom instruction and only self study in distance learning):

First method: 94 88 91 74 87 97

Second method: 85 82 79 84 61 72 80

Third method: 89 67 72 76 69

Use the H test at the 0.05 level of significance to test the null hypothesis that the three methods are equally effective.
(10 marks)

QUESTION 4:

The green pod yield (kg) under four treatments is as tabulated below

No. of plots	Treatment			
	1	2	3	4
1	3.17	3.44	3.15	2.48
2	3.40	2.88	2.69	2.37
3	3.50	2.97	3.10	2.58
4	2.87	3.27	2.80	2.84
5	3.88	3.94	3.45	3.00
6	4.00	3.87		2.48
7	3.60	3.25		

The hypothesis that there is no difference among four treatments by:

i) the median test (10 marks)
ii) the Kruskal walli's test (10 marks)

QUESTION 5:

(a)For a given set of paired data, $\{(x_1, y_1): i=1, 2, 3 \ldots n\}$ define Spearman's rank correlation coefficient,,r_s.

(b)Twelve sets of identical twins were given psychological tests to measure their aggressiveness. The emphasis was on the examination of the degree of similarity between twins within the same set.

The data were measures of aggressiveness and are given below:

Twin set	1	2	3	4	5	6	7	8	9	10	11	12
1^{st} born xi	86	71	77	68	91	72	77	91	70	71	88	87
2^{nd} born yi	88	77	76	64	96	72	65	90	65	80	81	72

(i)Calculate the rank correlation coefficient r_s for these data set.

(ii)Test at $\alpha=0.01$ level of significance whether the value of r_s so obtained in 5b (i) above is statistically significant.

(iii)Comment on the relationship between the twins in terms of aggressiveness.