(Knowledge for Development)

KIBABII UNIVERSITY

UNIVERSITY EXAMINATIONS

2021/2022 ACADEMIC YEAR

FIRST YEAR SECOND SEMESTER

MAIN EXAMINATION

FOR THE DEGREE OF MASTER OF SCIENCE IN STATISTICS

COURSE CODE: **STA 814**

COURSE TITLE: THEORY OF NON-PARAMETRIC STATISTICS

DATE: 05/10/2022 **TIME:** 2 PM -4 PM

INSTRUCTIONS TO CANDIDATES

Answer Question One and Any other two

TIME: 2 Hours

This Paper Consists of 4 Printed Pages. Please Turn Over.

QUESTION ONE (30 MARKS)

(a). Differentiate between parametric and non-parametric statistical inference approaches. (6 marks)

b). The observed values of a random sample of size k from a continuous distribution are

10.2 14.1 9.2 11.3 7.2 9.8 6.5 11.8 8.7 10.8

Test $H_o: M = 8$ Vs $H_1: M > 8$ at 5% level of significance using the (8 marks)

i. Sign test

ii. Signed rank test.

c).Let $X_1, X_2, \ldots X_n$ denote a random sample from a continuous population with cumulative distribution function, $F_X(x)$. Show that if $X_{(1)} < \ldots < X_{(n)}$ denote the n order statistics from the population then the joint probability density function of the n order statistics

$$f(y_1, y_2, \ldots y_n) = n! \pi_{j=1}^{n} f_x(y_j) \quad , \quad 0 < y_1 < \ldots < y_n$$

(10 marks)

.where $f_x(x) \frac{d}{dx} F_x(x)$

(d). State and prove the probability integral transform theorem. (6 marks)

QUESTION TWO (20 MARKS)

a) Consider the following data on an empirical distribution function $S_n(x)$ and a hypothesized distribution $F_0(x)$ on 25 subjects and selected values of a random variable X.

X=x	1	4	10	25	60	80	100
$nS_n(x)$	4	10	13	17	21	24	25
$nF_o(x)$	2	5	9	16	17	19	25

Use the Kolmogorov-Smirnov two sided test procedure to test at 5% level of significance the hypothesis H_o $F(x) = F_o(x)$ against $H_a: F(x) \neq F_o(x)$

b) Suppose $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ is a random sample from a Continuous bivariate distribution with the distribution function F(x,y). Let $D_i = Y_i - X_i$; $i = 1, 2, \ldots, n$. Assuming that $D_1, D_2, \ldots, D_n$ is a random sample of difference from F(x,y), construct a $100(1 - \alpha)\%$ confidence interval for the unknown median of differences M_D based on the statistic T^+.

QUESTION THREE (20 MARKS)

(a) In an effort to determine immunoglobulin D(igD)levels of a certain ethnic group ,a large num ber of the blood samples representing both sexes for 12 year olds were taken .The following sam ple data gives the igD levels (inmg/100ml)

Male	9.3	0.0	12.2	8.1	5.7	6.8	3.6	9.4	8.5	7.3	9.7		
Female	7.1	0.0	5.9	7.6	2.8	5.8	7.2	7.4	3.5	3.3	7.5	7.0	

Use the large sample Wilcoxon rank sum test with the significance level ά=0.01 to test the hypoth esis that there is no differences between the sexes in the median level of the igD. (8 marks)

(b) Find the significance difference between treatment A and B using Mann-Whitney statistics

Treatment A	3	4	2	6	2	5
Treatment B	9	7	5	10	6	8

(6 marks)

(c) (n=2) , uniform population . let X_1 and X_2 be independent U(0,1) random variables. Find the joint t probability density function of X_1=min(X_1,X_2) and X_2=Max(X_1,X_2). (6 marks)

QUESTION FOUR (20 MARKS)

(a). To test the null hypotheses H_o:M = M_o, where M is the median of some continuous distribution function F(x). Let $D_i = X_i - M_0$, i = 1, 2,…………,n where M_0 is a specified value and X_1,,,,,,,,,X_n is a random sample from the population. Consider the random variable.

$$T_n^+ = \sum_{i=1}^{N} z_i\{rank(|D_i|)\}$$

where $$z_i = \begin{cases} 1, & if\ D_i > 0 \\ 0, & if\ D_i < 0 \end{cases}$$

(i). Derive the exact probability mass function (pmf) of T_4^+,

And using the pmf, obtain

(ii). Var (T_4^+),

(iii). Exact α = 0.125 level of significance for testing H_0 against H_a: not H_0

(15 marks)

(b). Consider the random variables X and Y with median m_x and m_y respectively. If $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ are random samples from X and Y respectively, explain how you would use the ordinary sign test to test

$H_0: M_x = M_y$

v/s $H_a: M_x \neq M_y$

Assume there is a possibility of "zeros". (9 marks)

QUESTION FIVE (20 MARKS)

a) Define order statistics (3 marks)

b) Explain the importance of order statistics (6 marks)

c) Let X_1, X_2, X_3, X_4 and X_5 be a random sample from a population from a distribution with pdf

$$f(x) = 2x \qquad 0 \leq x \leq 1$$

Let $Y_1 \leq Y_2 \leq Y_3 \leq Y_4 \leq Y_5$ denote the associated order statistics

(i) find $P(Y_4 < \frac{1}{2})$ (4 marks)

(ii) find the CDF of $Y_4 = P(Y_4 \leq y) \quad 0 \leq y \leq 1)$ (4 marks)

(iii) find the pdf of Y_4 (3 marks)