(Knowledge for Development)

KIBABII UNIVERSITY
UNIVERSITY EXAMINATIONS
2020/2021 ACADEMIC YEAR

MAIN CAMPUS

MASTERS FIRST YEAR SECOND SEMESTER EXAMINATIONS

COURSE CODE: STA 831

COURSE TITLE: NON-PARAMETRIC METHODS

DATE: 19/5/2021 **TIME:** 9:00 A.M – 12:00 NOON

INSTRUCTIONS TO CANDIDATES:

Answer Question one and any other two question

QUESTION ONE (30MARKS)

(a). Explain the meaning of the following as applied in non-parametric statistical inference.

i). Non-parametric hypothesis

ii). Distribution-free statistics

iii). m-sample tests

iv). the Kruskal-Wallis statistic. **(8 marks**

(b). Let $X_1, X_2, \ldots\ldots\ldots X_n$ denote a random sample from a continuous population with cumulative distribution function, $F_X(x)$. Show that if

$X_{(1)} < \ldots\ldots\ldots\ldots < X_{(n)}$ denote the n order statistics from the

Population then the joint probability density function of the n order statistics

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

.where

$$f_x(x) = {}^{d}/_{dx} F_x(x)$$ (13 marks)

c). 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 $H_a{:}F(x) \neq F_o(x)$ **(8 marks)**

QUESTION TWO (20 MARKS)

a). State and prove the probability integral transform theorem

b). Show that if U(r) is the r^{th} order statistic from the uniform distribution on the interval (0,1), for r= 1,2,…,n then

$$E\left(U_{(r)}^{k}\right)=\frac{\alpha(\alpha+1)\ldots\ldots\ldots(\alpha+k-1)}{(\alpha+\beta)(\alpha+\beta+1)\ldots\ldots(\alpha+\beta+k-1)}$$

where $\alpha = r-1$ and $\beta = n-r+1$ **(6 marks)**

c). Using the results in part (b), or otherwise determine

i). E(U(r))

ii). Var ($U_{(r)}$) **(6 marks)**

d).Show that if $X_{(r)}$ is the r-th order statistic from a continuous population with cumulative distribution function, F(x) , then $E\left(U_{(r)}^{k}\right) \cong F^{-1}\left(\frac{r}{n+1}\right)$ **(6 marks)**

QUESTION THREE (20 MARKS)

(a). Describe the Karl-Pearson goodness-of –fit test statistics and show that under appropriate conditions(specify) it has a chi-square distribution. (6 marks)

(b). Let $X_1, X_2, \ldots\ldots X_n$ be a random sample from some continuous distribution, F(x).

(i). Define the corresponding empirical distribution function $S_n(x)$

(ii). Show that $S_n(x)$ is an unbiased consistent estimator of F(x). (8 marks)

(c). Let $D_n^{+} = \sup_{x} \{S_n(X) - F(X)\}$

Given for y $\geq$ 0

$$\lim_{n \to 0} P\left(D_n^{+} < y/\sqrt{n}\right) = 1 - e^{-2y^2}$$

Deduce the limiting distribution of the statistics $V = 4n(D_n^{+})^2$(6 marks)

QUESTION FOUR (20 MARKS)

(a). Let M_x denote the median of the distribution for the random variable X.

Suppose $X_1\ X_2 \ldots\ldots\ldots X_n$ is a random sample from the distribution.

To test

$H_0: M_x = M_0$, M_o known verses $H_a : M_x \neq M_0$, the Wilcoxon's signed rank test st atistic, T^+, may be used.

(i). Define T^+

Show that when H_0 is true, for a random sample of size n,

(ii). $E(T^+) = \frac{n(n+1)}{4}$

(iii). $Var(T^+) = \frac{n}{24}(n - 1)(2n+1)$ (12 marks)

(b). Suppose $(X_1, Y_1),(X_2, Y_2),\ldots\ldots,(X_n, Y_n)$is a random samp le from a
Continuous bivariate distribution with the distribution function F(x,y).
Let $D_i = Y_i - X_i$;i = 1, 2,………………………..,n
Assuming that $D_1, D_2, \ldots\ldots, D_n$ is a random sample of diff erence
from F(x,y), construct a $100(1 - \alpha)\%$ confidence interval for the unknown median of differences M_D based on the statistic T^+. (8 marks)

QUESTION FIVE (20 MARKS)

a) Explain the advantages of Non-parametric statistical inferences over the Traditional parametric inferences. (10 marks)

(b) A researcher planted maize at the same rate in 8 small plots of ground, then weeded the maize rows by hand to allow no weeds in 4 randomly selected plots and exactly 3 lamb's-quarter weed plants per meter of row in the other 4 plots. The table below gives data on the yield of maize per acre in each of the plots.

Weeds per meter	**Yield**			
0	166.7	172.2	165.0	176.9
3	158.6	176.4	153.1	156.0

Test
H_0: No difference in distribution of yields
Versus
H_a: Yields are systematically higher in weed- free plots
Using:

i) W_N

ii) H **(10 marks)**