(Knowledge for Development)

KIBABII UNIVERSITY

UNIVERSITY EXAMINATIONS

2017/2018 ACADEMIC YEAR

FIRST YEAR SECOND SEMESTER

SPECIAL/SUPPLEMENTARY EXAMINATION

FOR THE DEGREE OF MASTER OF SCIENCE IN STATISTICS

COURSE CODE: STA 831

COURSE TITLE: THEORY OF NONPARAMETRIC STATISTICS I

DATE: 08/10/18 **TIME:** 8 AM - 11 AM

INSTRUCTIONS TO CANDIDATES

Answer Question ONE and any Other TWO Questions

TIME: 3 Hours

Q1). 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). Problem of zeros in sign tests and their solution. **(8 marks)**

b). Let $S_n(x)$ denote the empirical distribution function for a continuous population with cumulative distribution function, F(x)

i). Show that $S_n(x)$ is unbiased consistent estimator of F(x)

ii). Define using $S_n(x)$, the Kolmogorov-Smirnov statistics D_n, D_n^+ and D_n^-

iii). Describe, with respect to the statistics stated in part b (ii), the corresponding null and alternative hypotheses. **(12 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)**

Q2. a). i). Define the Kruskal Wallis Statistic, H, for testing the null hypothesis

H_0: F_{x_1} (x) = F_{x_2} (x) = ... F_{x_m} *(x)*

Versus

H_a : NOT H_0

where F_{x_j} (x) is the cumulative distribution function of population X_j

for j=1, 2...., m

ii) For m=2, state the corresponding Wilcoxon Rank Sum test statistic

W_N **(8 marks)**

b) State the distribution of the statistic

i)H

ii)W_N **(4 marks)**

c) 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 **(8 marks)**

Q3 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 α= r-1 and β= 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)

Q4. Consider the Wilcoxon's Signed Rank test statistic T_N^+ for testing

H_o : M = M_0

v/s

H_a:M $\neq$ M_o

where M is the population median and M_o is some known constant.

a). Define T_N^+ **(5 marks)**

b). Compute the exact probabilities;

P($T_8^+ \geq$ j)

for j = 33, 34, 35, 36 **(7 marks)**

c). Using the results in part (b) or otherwise, determine the exact level of significance, α, for each of the following critical regions of the test.

i). Critical region = $\{T_8^+ : T_8^+ \leq 2 \; or \; T_{10}^+ \geq 34\}$

ii). Critical region $\{T_8^+ : T_8^+ = \leq 3 \; or \; T_{10}^+ \geq 33\}$ **(8 marks)**

Q5. a). Let X_1, X_2,..........,X_n be a random sample from a population with cumulative distribution function denoted by F(x).

i). Show that the confidence interval for the p-th percentile θ_p for some given confidence coefficient $\alpha, 0 < \alpha < 1$, is of the form ($X_{(r)}$, $X_{(s)}$), r < s where $X_{(j)}$ is the j-th order statistic from the population,

j = 1,2,........,n

ii). Describe how the integers r and s can be determined. **(10 marks)**

b). Let $U_{(r)}$ denote the r-th order statistic, r = 1,2,......,n from a uniform distribution on the interval (0, 1). Let $Z_{(r)}$ = $(U_{(r)} - \mu)/_\sigma$ where μ = E($U_{(r)}$) and $\sigma^2 = Var(\text{U(r)})$. Show that $Z_{(r)} \xrightarrow{P} Z$ where Z has N(0, 1) distribution.

(10 marks)