KISII UNIVERSITY
MAIN CAMPUS
UNIVERSITY EXAMINATIONS
FIRST YEAR EXAMINATIONS FOR THE DEGREE OF MASTERS OF MATHEMATICAL STATISTICS
SECOND SEMESTER 2022/2023
(JANUARY-APRIL 2023)

MATH 876: NONPARAMETRIC INFERENCE

STREAM: MSc Y1S2 **TIME**: 2 HOURS
DAY: **DATE**:

INSTRUCTIONS:

1. *Do not write anything on this question paper.*
2. *Answer question **ONE** and any other **TWO** questions.*

Question One (Compulsory) (30 Marks)

(a) Explain clearly the following concepts of nonparametric inference.

(a) Order statistics. (5 marks)

(b) Distribution free statistics. (4 marks)

(b) Let $X_1, \cdots, X_n$ be iid absolutely continuous random variables with distribution function $F(x)$ and $X_{(1)} < \cdots < X_{(n)}$ their order statistic. Derive

(i) The joint density function of the order statistics of $X_1, \cdots, X_n$. (5 marks)

(ii) The probability density function of the i^{th} order statistics. (4 marks)

(c) Carefully state and prove the probability integral transformation (PIT) and briefly explain its importance in nonparametric inference. (7 marks)

(d) Find the quantile function, $Q(p)$ for a random variable X from the exponential distribution with $\beta = 2$ and CDF (5 marks)

$$F(x) = \begin{cases} 0, & x < 0 \\ 1 - e^{-\frac{x}{2}}, & x \geq 0 \end{cases}$$

Question Two (20 Marks)

(a) With the aid of some real life examples, explain the circumstance(s) under which the chi-square goodness of fit(GoF) test may be used and suggest alternative tests to the chi-square GoF test (3 marks)

(b) Show that the chi-square GoF test is a consistent test and hence, also that it is asymptotically unbiased.

Question Three (20 Marks)

(a) Write the k-sample median test statistic, where the null hypothesis is that $k(\geq 2)$ populations are identical against the general alternative that the populations are different in some way.

(b) For the arbitrary data below, use the median test to test whether there exist any difference between the three groups

Group	1	2	3
	73	96	12
	79	92	26
	86	89	33
	91	95	8
	35	76	78

Question Four (20 Marks)

(a) Distinguish between cumulative distribution function $F(x)$ and the empirical distribution function $S_n(x)$.

(b) The Kolmogorov-Smirnov statistics are defined as

$$D_n =_{x\in\Re}^{sup} \Big[S_n(x) - F(x)\Big], \quad D_n^- =_{x\in\Re}^{sup} \Big[F(x) - S_n(x)\Big], \quad D_n^+ =_{x\in\Re}^{sup} \Big[S_n(x) - F(x)\Big]$$

where $F(x)$ is a continuous cumulative distribution function and $S_n(x)$ is the empirical distribution function.

(i) Prove that the statistics D_n, D_n^- and D_n^+ are completely distribution free.

(ii) Explain the advantage of the the three statistics being distribution free.

Question Five (20 Marks)

Let $X_1, \cdots, X_n$ and $Y_1, \cdots, Y_m$ denote iid real-valued random variables with continuous distribution functions $F(x)$ and $G(x) = F(x - \triangle)$, respectively, where $\triangle \in \Re$ is an unknown shift parameter. Let $R^* = (Q_1, \cdots, Q_m, R_1, \cdots, R_n)$ denote the vector of ranks for

$(X_1, \cdots, X_m, Y_1 \cdots, Y_n)$. Then

$$W = \sum_{n}^{i=1} R_i \quad U = \sum_{i=1}^{m} \sum_{j=1}^{n} \Psi(Y_j - X_i)$$

are called the Wilcoxon and the Mann-Whitney rank sum statistics, respectively. Under the null hypothesis $H_0 : \triangle = 0$, W is a rank statistic. Show that $W = U + \frac{n(n+1)}{2}$ implying that under $H_0 : \triangle = 0$, U is a rank statistic too.