(Knowledge for Development)

# KIBABII UNIVERSITY

## UNIVERSITY EXAMINATIONS

## 2016/2017 ACADEMIC YEAR

## FOURTH YEAR FIRST SEMESTER

## SPECIAL/ SUPPLEMENTARY EXAMINATION

## FOR THE DEGREE OF BACHELOR OF SCIENCE

**MATHEMATICS**

**COURSE CODE:** **STA 449**

**COURSE TITLE:** **NON-PARAMETRIC METHODS**

**DATE:** 11/09/17 **TIME:** **3 PM -5 PM**

---

### INSTRUCTIONS TO CANDIDATES

Answer Question One and Any other TWO Questions

TIME: 2 Hours

## Question 1: (30 marks)

a) Let $X_1, X_{2,} \dots\dots X_m$ be $m$ independent observation from distribution of $F(x)$. Let also be $n$ independent observation from distribution of $G(y)$. If $T$ is the sum for ranks of $Y's$, show that the Wilcoxon's Rank sum statistic is $T \sim N(\frac{n}{2}(m+n+n), \frac{mn}{12}(m+n+1)$.

**[4mks]**

b) Marks obtained by 10 students in chemistry and maths are given below.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Chemistry | 84 | 74 | 48 | 54 | 72 | 71 | 96 | 75 | 69 | 100 |
| Maths | 69 | 64 | 66 | 72 | 85 | 68 | 87 | 86 | 71 | 91 |

Test using Rank correlation test the hypothesis that the two categories are independent. Use $\alpha = 0.05$ **[6mks]**

c) Suppose that the light of a certain bulb is exponentially distributed with mean 100 hours $F(x) = \frac{1}{100} e^{-\frac{1}{100}x} \quad x \geq 0$

If ten such light bulb are installed simultaneously, what is the distribution of the life of the life of the bulb that fails first and what is its expected life. **[5mks]**

d) For a sample of size 10. What is the confidence coefficient of the interval $X_{(2),} X_{(1)}$ which is the confidence interval estimator of the population media? **[2mks]**

e) Let $X_1, X_{2,} \dots X_n$ be a random sample of size n from a population whose continuous distribution is $F(x)$. If $P_1, P_{2,} \dots\dots P_k$ are the probabilities than an observation belong to the i[th] category i.e $P_i = P(x \epsilon A_i)$, where $\sum_{i-1}^{k} p_i = 1$ and that $\pi_i = P(x \epsilon A_i/H_0)$. Give the procedure for testing the chi-square goodness of fit test by using the likelihood function and hence show the karl-pearson $\chi^2$ test statistic. **[8mks]**

f) A genetic theory indicates that for a certain species of flowers, White, red and blue flowers should occur in the ratio 5:3:1. Suppose that in a random sample of 180 flowers 90 are white, 65 are red and 25 are blue. What frequencies would we expect if the theory is correct?

At 1% level of significance tests the genetic theory that:

$H_0$ : the genetic theory is correct

$H_1$ : the genetic theory is incorrect **[5mks]**

## Question 2: [20 marks]

a) Twenty-three applicants for a position are interviewed by three administrators and rated on a scale of 5 as to suitably for the position. Each applicant is given a 'suitability' score which is the sum of the three numbers. Although college education is not a requirement for the position, a personnel director felt that it might have some bearing on suitability for the position. Raters made their ratings on the basis of individual interviews and were not told the education background of the applicants. Twelve of the applicants had completed at least two years of college.

Use the mann-Whitney u-test to determine whether there was a difference in the scores of the two groups. Use a 0.05 level of significance.

Group A had an educational background of less than two years of college, while group B had completed at least two years of college.

"Suitability" scores

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Group A | 7 | 11 | 9 | 4 | 8 | 6 | 12 | 11 | 9 | 10 | 11 | 11 |
| Group B | 8 | 9 | 13 | 14 | 11 | 10 | 12 | 14 | 13 | 9 | 10 | 8 |

**[8mks]**

b) A die is thrown 120 times with the following results.

| | | | | | | |
|---|---|---|---|---|---|---|
| Face | 1 | 2 | 3 | 4 | 5 | 6 |
| Frequency | 18 | 23 | 16 | 21 | 18 | 24 |

Is the die fair?

Test at $\alpha = 0.05$ level of significance. **[5mks]**

c) The number of automobiles declared abandoned and disposed off in a town in 104 weeks is given by the following:

| | | | | |
|---|---|---|---|---|
| No. of automobiles | 0 | 1 | 2 | 4 or more |
| No. of weeks | 40 | 43 | 17 | 4 |

Test the hypothesis that the number of cars declared abandoned per week is a Poisson random variable with parameter $\lambda = 1$ with $\alpha = 0.05$ **[7mks]**

## Question 3: [20 marks]

a) The green pod yield (kg) under four treatments is 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 | | |

Test the hypothesis that there is no difference among four treatments using the Kruskal-Wall's test. Use $\alpha = 0.05$ **[6mks]**

b) The iron determination in five pea-leaf sample each under three treatments, were as tabulated below.

| Sample number Blocks | Treatments | | |
|---|---|---|---|
| | 1 | 2 | 3 |
| 1 | 591 | 682 | 727 |
| 2 | 818 | 591 | 863 |
| 3 | 682 | 636 | 773 |
| 4 | 499 | 625 | 909 |
| 5 | 648 | 863 | 818 |

Apply Friedman's test to confirm whether

$H_0$: the iron content in leaves under treatment is the same.

$H_1$: that at least two of them have different effects. (Use $= 0.05$ ) **[4mks]**

g) Define an order statistics **[3mks]**

h) Let $X_1, X_{2,} \dots \dots\ X_n$ be a random sample from a population whose cumulative distribution function is $F(x)$. Let $X_{(r)}$ denote the $r^{th}$order statistic and $F_r(x) = P(X_{(r)} \leq x)$ be the cdf of $X_{(r)}$ . Obtain the probability distribution function (pdf) of order statistic when $r = 1$ and $r = n$. **[7mks]**

## Question 4: [20 marks]

a) Consider a statistical experiment with $k$ possible outcomes $A_{1,}\quad A_2$ ,..... $A_k$ such that $P(A_i) = P_i$ and$\sum_{i-1}^{k} p_i = 1$. Suppose the experiment is carried out in n independent times. Let denote the number of occurrence (outcome) of $n$ which belong to $A_i$. If the joined distribution of $X_1, X_{2,} \dots\ X_k$ follows a multinomial distribution given by $P(X_1 = x_1, \quad X_2 = x_2, \cdots, X_k = x_k) = f(x_1, x_2, \cdots, x_n)$. Show that the pdf of the $r^{th}$order statistic is given by $f_r(x) = \frac{n!}{(r-1)!(n-r)!}[F(x)^{r-1}][1 - F(x)^{n-r}]\,f(x)$ . **[6mks]**

b) Two methods of instructions were administered to 10 pairs of students taking a course in statistic. A student was elected at random from each pair and method A administered to him or her and the remaining student underwent method B. Then after the instructions a test was given to both groups of students with the following scores

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| A | 65 | 40 | 59 | 71 | 35 | 62 | 80 | 75 | 43 | 56 |
| B | 49 | 60 | 81 | 78 | 90 | 48 | 64 | 54 | 72 | 50 |

Using a median test, test whether the methods of instructions are equally effective against A is better than B. (use $\alpha = 5\%$) **[8mks]**

c) A typing school claims that in a six week intensive course, it can train students to type, on the average at least 60 words per minute. A random sample of 15 graduates is given a typing test and the median number of words per minute typed by each of these students is given below.

| Student | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Word per min | 81 | 76 | 53 | 71 | 66 | 59 | 88 | 73 | 80 | 66 | 58 | 70 | 60 | 56 | 55 |

Test hypothesis that the median typing speed of graduates is at least 60 words per minute. (use $\alpha = 5\%$) **[6mks]**

## Question 5: [20 marks]

a) Show that in Asymptotic distribution of $R$ we have $R \sim N\ 2n\lambda(1-\lambda), 4n\lambda^2(1-\lambda)^2)$ where we assume that $n \to \infty$ in such a way that $\lambda = \frac{n_1}{n}$ and $1 - \lambda = \frac{n_2}{n}$ remain constant. **[8mks]**

b) Given the following data

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Before | 30 | 28 | 34 | 35 | 40 | 42 | 33 | 38 | 34 | 45 | 28 | 27 | 25 | 41 | 36 |
| After | 32 | 29 | 33 | 32 | 37 | 43 | 40 | 41 | 37 | 44 | 27 | 33 | 30 | 38 | 36 |

Use the sigh test to see if there is a difference between the number of days until collection of an account receivable before and after a new collection policy. Use the 0.05 level of significance.

**[5mks]**

c) It is claimed that successive that last digit of telephone number in an alphabetical listing by name constitute a set of random numbers opening a telephone directory at random and starting with a random selected print. A sequence of 20 last digit is observed since the last digit is equally likely an integer between 0-9 inclusive. The population median is 4.5 the number obtained were
Last digit: 7, 0, 4, 8, 5, 3, 1, 5, 0, 2, 6, 8, 0, 9, 9, 4, 0, 1, 8, 1 for large n.

Test $H_0$: the twenty numbers appear I a random order.
V/s $H_1$: the twenty numbers do not appear in a random order. (use $\alpha = 0.05$) **[7mks]**