KENYATTA UNIVERSITY

UNIVERSITY EXAMINATIONS 2011/2012

SECOND SEMESTER EXAMINATION FOR THE DEGREE OF MASTER OF SCIENCE

SMA 467: NON-PARAMETRIC METHODS

DATE: Wednesday, 4th April, 2012 TIME: 8.00 a.m. – 10.00 a.m.

INSTRUCTIONS: Answer question **ONE** and any other **TWO** questions.

Q1 (a) Distinguish between parametric and non-parametric tests (2marks)

(b) Explain the concept of runs and how they could be used for testing randomness. (3marks)

(c) What are order statistics? Are they independent? Explain. (3marks)

(d) Let $X_1, X_2, \ldots, X_n$ be a random sample from a population whose density function is

$$f(x) = \begin{cases} e^{-x} & ,x > 0 \\ 0 & ,elsewhere \end{cases},$$

Let $Y_r = X_{(r)}$ be the r-th order statistics.

Find (i) the joint distribution of $Y_1, Y_2, \ldots\ldots Y_n$ (3marks)

(ii) marginal distribution of Y_1 (5marks)

(iii) median of the distribution (3marks)

(e) Ten students were tested before and after being coached in a subject. Their marks were as follows:

Student number	1	2	3	4	5	6	7	8	9	10
Before coaching	53	59	61	48	39	56	75	45	81	60
After coaching	60	57	67	52	61	71	70	46	93	75

Test, at the 5% level of significance, the hypothesis that coaching has no effect on average marks against a two sided alternative hypothesis using a sign test. (5marks)

(f) What is a goodness of fit test? (2marks)

(g) Suppose that four minutes is claimed to be the median taken by a factory's machine operators to complete a particular task that they carry out. Many points in time, an operator was observed and the time taken to complete the task being

3.5, 3.2, 3.9, 4.0, 4.8, 3.8 and 3.4

Do these figures cast doubt on the claimed median time? Use an appropriate Non-parametric test at 10% level of significance (4marks)

Q2 (a) Consider an ordered sequence of two symbols of types I and II. Let R_1 and R_2, n_1 and n_2 be runs and number of symbols of type I and II respectively. Write down the mean and variance of R and hence obtain the asymptotic distribution of R. (7marks)

(b) Items emerging from a continuous production line were classified as defective, D or non-defective, N. A sequence of items over time was as follows:

DNNNDDNDNNNDDDNNDDNDND

Using large approximation for the run test, test whether these data suggest lack of randomness in occurrence of defective and non-defective. Use α=0.05 (6marks)

(c) The winning time for the Olympic Championships in the 1500 meter race for men has been recorded for each Olympic game since 1896. For the period 1896-1972, the record of times in minutes and seconds is shown below.

4:33.2 , 4:06.0 , 4:05.2 , 4:03.2 , 3:56.8 , 4:12.0 , 4:01.4 , 3:53.6 , 3:53.2, 3:51.2 , 3:47.8 , 3:49.8 , 3:45.2 , 3:35.6 , 3:41.2 , 3:38.1 , 3:34.9 , 3:36.3

Use an appropriate run test to analyze these time ordered observations for randomness against an alternative of insufficient fluctuation. Comment on your conclusions. Use α=0.05 (7 marks)

Q3. (a) Let $X_1, X_2, \ldots, X_n$ be independent observations from a continuous distribution F(x). Let $F_o(x)$ be a completely known distribution.

Derive a chi-square test of goodness fit procedure to test the hypothesis

$H_o : F(x) = F_o(x)$ against

$H_1 : F(x) \neq F_o(x)$

(13 marks)

(b) A theory predicts that the proportion of beans into four groups A, B, C and D should be in the ratio 9 : 3 : 3 : 1. In an experiment among 1600 beans, the numbers in the four groups were 882, 313, 287 and 118 respectively. Does the experimental result support the theory. Use $\alpha = 0.05$ (7marks)

Q4 (a) Let $X_1, X_2, \ldots, X_m$ be m independent observations from a continuous distribution F(x). Let $Y_1, Y_2, \ldots, Y_n$ be n independent observations from a continuous distribution G(y). Derive the Mann-Whitney U statistic in testing whether the distributions for X and Y are identical. Let T be the sum of ranks of Y's in the combined ordered sample of X's and Y's. Derive the Wilcoxon's rank sum criterion for testing

$$H_0 : F(x) = G(y) \quad \text{against}$$

$$H_1 : F(x) \neq G(y).$$

Write down the approximate distribution of U under H_0. (13marks)

(b)An experiment was conducted to compare the strength of two types of papers; one standard and the other treated. Seven pieces of each type of paper were randomly selected and the strength measurement noted as shown below.

Standard (X)	1.21	1.43	1.35	1.51	1.37	1.17	1.48
Treated (Y)	1.49	1.37	1.67	1.50	1.31	1.29	1.52

Use the Wilcoxon's test statistic to test the hypothesis of no difference in the distributions of the strengths for the two types against the alternative that the treated paper tends to be of greater strength. Use 5% level of significance (7marks)

Q5 (a) Consider two random variables X and Y with joint continuous distribution F(x,y) and the hypothesis

H_0: X and Y are uncorrelated against

H_1: X and Y are correlated

Write down the Spearman's rank correlation coefficient of X and show how you can test H_0 against H_1. (5marks)

(b) Sixteen subjects were matched in pairs for their ability to perform a certain task. One subject in each pair was given three double whiskies and asked to perform the task again a few minutes later. The other subjects were completely sober when they performed the task again. The following scores were obtained

Pair	1	2	3	4	5	6	7	8
Sober group	340	290	270	370	330	310	320	300
Whiskey group	320	360	330	540	370	300	680	1180

Test at 5% level of significance whether sober and whiskey tasks are uncorrelated against that they are negatively correlated. (5marks)

(c) Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a continuous distribution F(x). Let $X_{(r)}$ denote the r-th order statistic .Obtain the probability density function $X_{(r)}$.

If the pdf of X is given by $f(x) = \frac{1}{100}e^{-\frac{1}{100}x}$, x>0

Find the pdf of the 10^{th} order statistic. (10marks)