



MACHAKOS UNIVERSITY

University Examinations 2021/2022 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

FOURTH YEAR SECOND SEMESTER EXAMINATION FOR
BACHELOR OF SCIENCE (STATISTICS AND PROGRAMMING)

BACHELOR OF SCIENCE (MATHEMATICS)

SST 403: NON-PARAMETRIC ESTIMATION

DATE: 7/12/2021

TIME: 8:30 – 10:30 AM

INSTRUCTION:

Answer Question One and Any Other Two Questions

QUESTION ONE (COMPULSORY) (30 MARKS)

- a) Distinguish between parametric and non-parametric statistical inference. (4 marks)
- b) Let X_1, X_2, X_3 be a random sample from a uniform distribution over the interval (0,1). Find the
- median of the distribution.
 - probability that the sample median is less than the median of the distribution.
 - mean of the sample median. (7 marks)
- c) The following are values of a random sample from a population having a continuous distribution with median M:
2.4, 2.8, 3.6, 4.1, 3.9, 2.2, 3.8
Test $H_0: M = 3$ vs $H_1: M > 3$
Using the
- Probability value.
 - Critical region and give exact level of significance of the sign test at 5% level of significance. (8 marks)

- d) Assume $X_1, X_2, \dots, X_{100}$ constitute a random sample from uniform distribution over $(0, 2)$. Determine the expectation and variance of the empirical distribution function at $x = 0.8$ i.e., $\hat{F}_n(0.8)$. (7 marks)
- e) Given the data $\chi = \{1, 2, 6, 11, 13, 14, 20, 23\}$. Determine the k-NN (k-nearest neighbor) density estimator at $x = 5$ with $k = 2$ and with $k = 5$. (4 marks)

QUESTION THREE (20 MARKS)

- a) A random sample of 200 people were asked the number of times they had watched football matches over a period at Nyayo Stadium and the following results were obtained.

No. of matches	No. of people
0	52
1	60
2	55
3	18
4	8
5 or more	7

Test the null hypothesis that the 200 observations form a random sample from the Poisson distribution at 5% level of significance. (14 marks)

- b) Ten Students were tested before and after being coached in a subject. Their marks were as follows:

Serial No.	1	2	3	4	5	6	7	8	9	10
Before	53	59	61	48	39	57	75	45	81	60
After	60	57	67	52	61	71	70	46	93	75

Use the sign test to test the hypothesis that coaching does not improve the performance of the students at 5% level of significance. Give the exact level of significance. (6 marks)

QUESTION TWO (20 MARKS)

- a) Consider the waiting times, in minutes of people at XYZ bank given below

Waiting time (x)	0-	0.5-	1.0-	1.5-2.5
No. of people	77	60	35	28

It is claimed that the waiting times can be modelled by the probability distribution

$$f(x) = \begin{cases} 0.8 - 0.32x, & 0 \leq x \leq 2.5 \\ 0, & \text{otherwise} \end{cases}$$

Test this claim at 10% significance level.

(8 marks)

- b) A simple study with only 2 subjects in each group yielded the below values

Treatment	175	250
Control	255	275

Use the permutation test to access the hypotheses H_0 : The treatment has no effect Versus H_1 :

The treatment has an effect at $\alpha = 0.05$ level of significance.

(12 marks)

QUESTION FOUR (20 MARKS)

- a) A random sample of size 3 is taken from the density $f(x) = \begin{cases} 2x, & 0 < x < 1 \\ 0, & \text{otherwise} \end{cases}$

Find the mean of the median

(5 marks)

- b) Four identical six-sided dice, each with faces marked 1 to 6 are rolled 200 times. At each rolling, a record is made of the number of dice whose score on the uppermost faces is even.

The results are as follows,

No. of scores	0	1	2	3	4
Frequency	10	41	70	57	22

Assuming that the dice are equally biased in some unknown way, determine the appropriate distribution to model the data hence perform a goodness of fit test at $\alpha = 0.05$ level of significance.

(10 marks)

- c) Prove the zero expected value (zero tolerance) property of an influence function using the mean of an arbitrary distribution $F(x)$.

(5 marks)

QUESTION FIVE (20 MARKS)

- a) Let $X_1, X_2, \dots, X_n$ be a random sample from the uniform distribution over the interval $(0, 1)$.

Let $Y_r = X_{(r)}, r = 1, 2, \dots, n$ be the r^{th} order statistic.

Let $U = \frac{Y_r}{Y_s}, r < s$ and $V = Y_s$.

Show that

- i. U has a beta distribution with parameters r and $s - r$.
 - ii. U and V are statistically independent. (12 marks)
- b) An experiment to investigate the effect of vaccination against a certain disease yielded the following results

	Got disease	Did not get disease
Vaccinated	9	42
Not vaccinated	17	28

- i. Test the hypothesis that there is no difference between the vaccinated and unvaccinated groups at 5% level of significance.
- ii. Find the contingency coefficient for the above data.
- iii. Comment on the result of (ii) with respect to maximum value of the contingency coefficient. (8 marks)