



MACHAKOS UNIVERSITY

University Examinations for 2020/2021 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

FOURTH YEAR FIRST SEMESTER EXAMINATION FOR

BACHELOR OF SCIENCE IN MATHEMATICS AND COMPUTER SCIENCE

BACHELOR OF SCIENCE IN STATISTICS AND PROGRAMMING

BACHELOR OF SCIENCE IN MATHEMATICS

BACHELOR OF EDUCATION

BACHELOR OF ARTS

SMA 467: TEST OF HYPOTHESIS II

DATE: 18/08/2021

TIME: 11:00 – 1:00 AM

Instructions to the Candidate:

1. Answer **Question 1** and any other **two** questions.
2. You need a Scientific Calculator and Statistical Tables for this paper.

QUESTION ONE (COMPULSORY) (30 MARKS)

- (a) Explain the term *hypothesis* as used in statistics, illustrating with *two* examples from real life situations: (4 marks)
- (b) Explain each of the following terms as used in statistical test of hypothesis:
- (i) test statistic;
 - (ii) confidence interval;
 - (iii) p-value.
 - (iv) critical values

(8 marks)

- (c) The following is a set of marks scored in Mathematics in an examination by a class comprising of 32 students from a certain high school.

50 44 25 28 69 58 42 81 16 20 08 84 28 26 46 66
35 24 32 36 60 48 45 86 30 17 34 34 85 40 54 20

Test the set of marks for randomness at the 5% level of significance. Median = 43.
(8 marks)

- (d) The following data set represents the yields of maize in thousand sacks per farm block recorded in two agricultural zones – Zone A and Zone B.

Zone A	15.2, 8.9, 9.3, 14.4, 15.6, 11.8, 16.3, 17.8
Zone B	11.5, 12.6, 19.4, 21.3, 32.5, 18.6, 17.0, 23.4, 29.6

Using the Mann-Whitney test, assess whether there is a statistically significant difference in the mean yields between the two agricultural zones. Use the 5% level of significance.
(10 marks)

QUESTION TWO (20 MARKS)

- (a) (i) State **two** advantages and **two** disadvantages of non-parametric tests of hypothesis.
(4 marks)
- (ii) Non-parametric tests are said to be *less powerful* compared to parametric tests. Justify this statement.
(4 marks)
- (b) Explain **two** situations in which the Kolmogorov-Smirnov test can be used as a technique in statistical test of hypothesis.
(4 marks)
- (c) The marks scored in two papers in an examination by a random sample of 12 students from a certain high school is as shown in the table below.

Student	A	B	C	D	E	F	G	H	J	K	L	M
Paper 1	12	16	34	48	49	43	54	18	32	15	36	52
Paper 2	43	25	40	50	10	24	30	33	53	42	28	20

Using the Wilcoxon signed rank test for paired data, test whether the two sets of marks are drawn from the same subject (population) at the 5% level of significance. (8 marks)

QUESTION THREE (20 MARKS)

Two land economists formulated two regression models for estimating the price of a plot of land based on a specified set of input variables. To evaluate the two models, data was collected on the actual prices of a random sample of 12 plots, and then two price estimates were derived for each plot using each model. The actual price and the two estimated prices were recorded in ten thousand Kenya shillings for each plot as shown below:

Plot	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10	P11	P12
Actual	15	19	30	18	62	43	78	68	18	55	66	17
Model 1	20	10	45	28	34	45	82	90	45	56	75	25
Model 2	42	19	14	30	40	54	44	48	30	32	45	38

- Compute the Spearman's rank correlation co-efficient between the actual price for a plot and the Model 1 estimate, and between the actual price for a plot and the Model 2 estimate. (10 marks)
- Recommend the most suitable model for estimating the price of a plot of land, justifying your choice based on the results obtained in (i) above. (2 marks)
- Test for significance each of the co-efficients of correlation obtained in (i) above at the 5% level of significance. (5 marks)
- Determine the 99% confidence interval for the population correlation co-efficient. (3 marks)

QUESTION FOUR (20 MARKS)

- An educational researcher claims that the pass rate for subjects in KCSE varies between subjects with the proportion of students who pass in some of the subjects is as shown below:

English	50%	Geography	75%
Mathematics	30%	Physics	20%
Kiswahili	40%	Biology	60%

A random sample of 800 KCSE candidates in a certain year was taken for the students who sat for *all the six* listed subjects, and the number of candidates who passed in the seven subjects is as shown below:

English	500	Geography	640
Mathematics	190	Physics	140
Kiswahili	310	Biology	420

Using the chi-square test for goodness of fit, evaluate whether this claim about KCSE pass rates is true or not at the 5% level of significance. Assume that the performance among students and between subjects is random and independent. (8 marks)

- (b) A crime researcher carried out a study on the relationship between the type of crime committed and the age group of the convict. A random sample of 400 convicted criminals was collected, and the data pertaining to the age of a convict and the type of crime was summarized in the contingency table below.

		Youth	Middle Age	Old Age
Crime	Robbery	78	96	41
	Fraud	30	54	21
	Murder	32	25	23

Using the chi-square test for independence, evaluate whether there is a relationship between the type of crime and the age group. Use the 5% level of significance. (12 marks)

QUESTION FIVE (20 MARKS)

- (a) In a certain mango farm, there are different varieties of mangoes planted randomly and grown under similar conditions. The number of fruits in thousands from three random samples of mango trees in the farm is as shown in the table below.

Sample 1	3.8	3.0	4.2	2.5	4.4
Sample 2	2.9	2.7	3.3	4.0	2.4
Sample 3	2.8	3.4	3.6	3.7	2.5

Using the Kruskal-Wallis test, evaluate whether the three samples of mangoes are drawn from the same variety or not, at the 5% level of significance. (8 marks)

- (b) Given a set of n bi-variate observations comprising pairs of variables (x_i, y_i) , derive the formula for the Spearman's rank co-efficient of correlation r_{xy} , and hence, show that

$$r_{xy} = 1 - \frac{6 \sum (R_x - R_y)^2}{n(n^2 - 1)} \quad (12 \text{ marks})$$