



MACHAKOS UNIVERSITY

University Examinations for 2022/2023 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 IN MATHEMATICS AND COMPUTER SCIENCE

BACHELOR OF SCIENCE IN STATISTICS AND PROGRAMMING

BACHELOR OF SCIENCE IN ACTUARIAL SCIENCE

BACHELOR OF SCIENCE IN MATHEMATICS

BACHELOR OF EDUCATION

BACHELOR OF ARTS

SMA 467 : TEST OF HYPOTHESIS II

Date: 10/12/2022

Time: 08:30 – 10:30

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

- (a) Different between a *one-tailed test* and a *two-tailed test* as used in statistical test of hypothesis. (4 marks)
- (b) Non-parametric tests are said to be *less powerful* compared to parametric tests.
 - (i) Justify this statement.
 - (ii) Explain the term *power of a test* as used in this context. (5 marks)
- (c) Explain each of the following terms as used in statistical test of hypothesis:
 - (i) test statistic;

- (ii) confidence interval;
 (iii) p-value. (6 marks)

- (d) 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 = 38.
(7 marks)

- (e) The following data set represents the marks scored in mathematics by a random sample of 15 boys and 12 girls from a certain high school. There is a theory that the performance in mathematics is not the same between boys and girls.

Boys	58, 68, 32, 64, 54, 44, 56, 62, 66, 60, 52, 55, 40, 75, 61
Girls	45, 72, 48, 55, 65, 35, 70, 55, 53, 74, 73, 67

Using the Mann-Whitney test, evaluate this theory by assessing whether there is a statistically significant difference in the performance between boys and girls. Use the 5% level of significance. (8 marks)

QUESTION TWO (20 MARKS)

- (a) Explain the term *hypothesis* as used in statistics, illustrating with *two* examples from real life situations: (4 marks)
- (b) Compare the terms *accuracy* and *parsimony* as used in model fitting and model selection in statistics. (4 marks)
- (c) Explain **two** situations in which the Kolmogorov-Smirnov test can be used as a technique in statistical test of hypothesis. (4 marks)
- (d) There is a popular belief that students who perform well in Mathematics tend to perform poorly in English, and vice versa. A random sample of 12 students was taken from a certain high school, and the marks scored in English and Mathematics is as shown below.

Student	A	B	C	D	E	F	G	H	J	K	L	M
English	45	67	50	79	36	55	47	25	30	56	35	57
Mathematics	32	57	58	65	46	61	37	30	21	45	37	45

Using the Wilcoxon signed rank test, evaluate this belief by testing whether there is a significant difference in the marks scored in the two subjects at the 5% level of significance. (8 marks)

QUESTION THREE (20 MARKS)

A telecommunications engineer has developed two types of devices for estimating long distances on the earth surface. One device uses terrestrial communication while the other uses satellite communication. In evaluating these two devices, 12 random distances were taken between various points, and the actual distance and the distance measured by each device recorded in kilometres as shown in the table below:

Distance	A	B	C	D	E	F	G	H	J	K	L	M
Actual	16	23	25	30	62	25	78	22	68	25	32	20
Terrestrial	36	48	26	64	40	65	48	45	72	32	42	25
Satellite	20	25	30	28	48	30	82	22	70	24	30	18

- (i) Compute the Spearman's rank correlation co-efficient between the actual distance and the distance for the terrestrial device, and between the actual distance and the satellite based device. (10 marks)
- (ii) Recommend the best device for estimating distance, justifying your choice based on the results obtained in (i) above. (2 marks)
- (iii) Test for significance each of the co-efficients of correlation obtained in (i) above at the 1% level of significance. (6 marks)
- (iv) Determine the 99% confidence interval for the population correlation co-efficient. (2 marks)

QUESTION FOUR (20 MARKS)

- (a) 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 as shown below:

English	50%	Geography	75%
Mathematics	25%	Biology	60%
Kiswahili	40%		

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

English	450	Geography	615
Mathematics	210	Biology	420
Kiswahili	305		

Using the chi-squared test for goodness of fit, evaluate whether this claim about KCSE pass rates is true or not at the 1% level of significance. Assume that the performance among the students and between the subjects is random and independent. (6 marks)

- (b) A study was conducted to analyse the relationship between smoking and lung cancer among adult males. A random sample of 270 adult males was collected, and the data was summarised in the contingency table shown below.

Risk factor	Cancer	No cancer	Total
Smoker	100	40	140
Non-smoker	50	80	130
Total	150	120	270

- (i) Using the chi-squared test for independence, evaluate whether there is a relationship between smoking and lung cancer. Use 5% level of significance. (7 marks)
- (ii) Determine the *odds ratio* between smoker and non-smoker, taking *non-smoker* as the reference category. Interpret the result. (4 marks)
- (iii) Determine the 95% confidence interval for the *odds ratio* obtained in (ii) above. (3 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

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