



MACHAKOS UNIVERSITY

University Examinations 2018/2019

SCHOOL OF PURE AND APPLIED SCIENCES
DEPARTMENT OF MATHEMATICS, STATISTICS AND ACTUARIAL SCIENCE
FOURTH YEAR SPECIAL/SUPPLEMENTARY 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
SMA 467 : TEST OF HYPOTHESIS II

DATE: 24/7/2019

TIME: 8:30 – 10:30 AM

INSTRUCTIONS:

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

1. (a) (i) Explain the term *test of hypothesis* as used in statistics. (2 marks)
(ii) Differentiate between a *two tailed test* and a *one tailed test* as used in statistical test of hypothesis. (4 marks)
(ii) Outline *four* conditions or assumptions which must be satisfied for the chi-square test to be valid in testing a statistical hypothesis. (4 marks)

- (b) The following data set represents the weight in kilograms of a group of patients who visited a certain clinic in a general hospital.

50	44	25	28	32	58	42	81	66	20	18	35
28	68	46	72	24	37	32	36	48	45	86	64

Test the set of weights for randomness at the 5% level of significance. Take: median = 43. (6 marks)

- (c) A certain theory from the police predicts that the proportion of stolen cars among some popular brands Toyota, Nissan, Mazda, Pajero, Honda are always in the ratio 10 : 8 : 4 : 2 : 1 respectively. In a given period of 5 years, records show that out of a total of 500 stolen cars for the five brands, the numbers of cars for the brands are 220, 150, 86, 36 and 8 respectively. Evaluate whether this observation supports the theory or not at 5% and 1% levels of significant. (6 marks)

- (d) The following data set represents the marks scored by a random sample of 14 male students and 12 female students who sat for a mathematics test in Masaku High School.

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

Using the Mann-Whitney U test, assess whether there is a statistically significant difference in the academic performance between the male and female students. Take level of significance $\alpha = 5\%$.

(8 marks)

2. (a) Highlight any *two* possible sources of statistical hypothesis. (2 marks)
- (b) Outline *two* advantages and *two* disadvantages of non-parametric tests of hypothesis. (8 marks)
- (c) A study was carried out to determine whether there is a difference in academic performance between Paper 1 and Paper 2 among the subjects tested in KCSE exams. A random sample of 12 candidates was selected in a certain subject and the marks scored were as shown in the table below.

Student	A	B	C	D	E	F	G	H	J	K	L	M
Paper 1	72	42	66	64	25	18	70	65	82	28	41	75
Paper 2	36	40	35	48	34	24	32	22	37	30	43	40

Using the Wilcoxon signed rank test for paired data, evaluate whether the two papers of the subject are the same. Use the 5% level of significance. (10 marks)

3. An engineering company 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 the devices, 12 random distances were taken between various points, and the actual distance and also the distance for each of the two devices 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

- (a) 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 distance for the satellite based device. (10 marks)
- (b) Recommend the most suitable device for estimating distance, justifying your choice based on the results obtained in (a) above. (3 marks)
- (c) Test for significance each of the two co-efficients of correlation obtained in (a) above. (7 marks)
4. (a) An educational researcher carried out a study on the relationship between the student academic performance in KCSE examination and the residential status of a student as boarder or day-scholar. A random sample of 800 students was selected, and the data pertaining to the academic performance and the residential status summarised in the contingency table below.

	Boarding	Day
Distinction	75	25
Credit	240	120
Pass	145	105
Fail	60	30

Using the chi-square test for independence, evaluate whether there is a relationship between the student academic performance in KCSE and the residential status. Test at the 5% level and the 1% level of significance. (10 marks)

- (b) A researcher in finance has formulated a statistical model used in estimating the number of people defaulting in loan repayments based on the amount of loan borrowed. He derived the linear least squares regression model $y = 8.625 + 1.858x$, where y is the number of people defaulting in loan repayments and x is the amount of loan borrowed by a client in million Kenya shillings. In order to evaluate the model, data was collected from a random sample of 250 loan defaulters, and their corresponding loan amounts and the actual number of people who defaulted in loan repayment recorded as shown below.

Loan Amount	0	1	2	3	4	5	6	7	8	9	10
No of Defaulters	8	10	12	15	25	20	28	30	26	18	14

Using the chi-square test for goodness of fit, evaluate the model on its suitability to estimate the number of people who defaulted in loan repayment. Test at the 5% and 1% levels of significance. (10 marks)

5. (a) In a certain bank, three different types of customers classified randomly under three samples. The deposit amounts s in million Kenya shillings from three random samples of customers are as shown in the table below.

Sample 1	3.8	3.0	4.2	2.5	
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 customers belong to the same income bracket or not at 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, and hence show that:

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

where r_{xy} is the Spearman's rank co-efficient of correlation between the variables x and y . (12 marks)