



MACHAKOS UNIVERSITY

University Examinations 2021/2022 Academic Year

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS AND STATISTICS

FIRST 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 (ARTS AND SCIENCE)

SMA 160: INTRODUCTION TO PROBABILITY AND STATISTICS

DATE: 6/3/2023

TIME: 2.00 – 4.00 PM

INSTRUCTIONS

1. Answer Question ONE and any other **two** questions;
2. You must have a Scientific calculator for this paper.

QUESTION ONE (30 MARKS)

- a) By use of sketch diagrams, differentiate between each of the following as used in Statistics:
 - i. Positive skewness and negative skewness;
 - ii. Positive correlation and negative correlation. (8 marks)
- b) Explain the term *co-efficient of determination* as used in statistics. (2 marks)
- c) The data given below represents the frequency distribution of the various shoe sizes purchased by 200 customers during the first quarter of the year 2018.

Size of shoes	4	5	6	7	8	9	10	11
No. of customers	4	15	35	52	48	32	12	2

Estimate by calculation each of the following measures about the distribution the sizes of shoes:

- i. Median size;
- ii. Mean size;
- iii. Standard deviation of the size. (8 marks)

- d) An agricultural scientist carried out a study on the relationship between maize crop yield and amount of fertilizer per unit area of land. A random sample of 12 maize plots of farms in different counties in Kenya was taken. The data given below shows the amount of fertilizer in bags per hectare plot of land and the amount of crop yield in sacks produced.

Farm	A	B	C	D	E	F	G	H	J	K	L	M
Fertilizer	15	12	08	10	16	10	18	09	24	14	15	12
Crop yield	55	45	40	50	80	40	65	35	75	70	55	45

- (i) Construct a scatter diagram to present this data. (6 marks)
- (ii) Interpret the scatter diagram constructed in (i) above. (2 marks)
- e) A committee of 6 members is to be formed from a group of 8 men and their wives. Determine the number of ways in which the committee can be constituted if a man and his wife cannot both serve in the same committee. (4 marks)

QUESTION TWO (20 MARKS)

- a) Measures of central tendency are dependent of origin while measures of deviation are independent of origin. Account for this statement. (4 marks)
- b) Explain the **four** measurement scales for statistical data as outlined by Stevens (1946) giving two examples in each case. (8 marks)
- c) The data given below shows the distribution of heights in cm of 200 male employment seekers who turned up for the police recruitment exercise in Nairobi County.

Height in cm	140 - 150	150 - 160	160 - 170	170 - 180	180 - 190	190 - 200	200 - 210
No. of recruits	5	35	a	60	40	b	3

The two unknowns a and b are numeric constants representing respective frequencies.

- i. Given that the total frequency is 200 and the mean of this distribution is 172.15 , determine the values of the constants a and b . (4 marks)
- ii. Nairobi County has been given a quota to employ only 60 recruits. The recruiting officials have decided that to take only the top 60 recruits based on height. To implement this criterion, determine the minimum height which must be set for a recruit to qualify to join the police. (3 marks)
- iii. The government has decided that some of those who are unsuccessful in the police recruitment will be absorbed in the National Youth Service (NYS). For recruitment into the NYS, an applicant must have a minimum height of 165 cm. Based on this

requirement alone, determine the percentage of the total 200 recruits who would be absorbed into the NYS. (3 marks)

QUESTION THREE (20 MARKS)

a) By considering the simple linear regression model $y_i = a + bx_i + e_i$ for $i = 1, 2, 3, \dots, n$, derive each of the following using the least squares technique:

- The normal equations;
- The regression co-efficient b and the regression intercept a , and hence show that

$$b = \frac{\sum xy - n \bar{x} \bar{y}}{\sum x^2 - n \bar{x}^2} \quad \text{and} \quad a = \bar{y} - b \bar{x} \quad (10 \text{ marks})$$

b) An engineering company has developed two types of devices for estimating long distances between two points on the earth surface. One device uses terrestrial communication while the other uses satellite communication. In evaluating them, 12 random distances were taken between various points and the actual distance and also the distance estimated by 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	15	19	18	30	62	30	78	23	68	55	30	17
Terrestrial	24	26	48	64	40	54	45	48	72	32	48	25
Satellite	20	25	28	23	48	64	82	28	90	65	45	10

- Compute the Spearman's rank correlation co-efficient between the actual distance and the estimated distance by terrestrial device, and between the actual distance and the estimated distance by satellite based device. (8 marks)
- Recommend the most suitable device for estimating distance, justifying your choice based on the results obtained in (a) above. (2 marks)

QUESTION FOUR (20 MARKS)

The Ministry of Lands and Housing is interested in investigating the relationship between the monthly rent of a house and the distance from the nearest tarmac road in Machakos County. A survey was carried out in which a random sample of 12 typical two bed-roomed self contained residential houses in Machakos County was taken. The data relating the distance from the nearest tarmac road in hundred metres and the monthly rent paid for a house in thousand Kenya shillings is as shown below.

House	A	B	C	D	E	F	G	H	J	K	L	M
-------	---	---	---	---	---	---	---	---	---	---	---	---

Distance	24.6	18.8	15.2	35.5	18.2	32.0	17.4	26.2	45.4	33.8	21.5	12.0
Rent	12	32	24	15	30	16	35	10	08	12	18	34

- a) Compute the Pearson's product moment co-efficient of correlation between the distance from the tarmac road and the monthly rent paid for the house. (8 marks)
- b)
 - (i) Determine the least squares regression line of the monthly rent paid on the distance from the nearest tarmac road. (4 marks)
 - (ii) Interpret the regression line obtained in (b) (i) above. (4 marks)
 - (iii) Using the regression line obtained in (b) (i), estimate by calculation the monthly rent for a typical two bed-roomed house which is 2500 metres from the nearest tarmac road. (2 marks)
- c)
 - (i) Compute the co-efficient of determination between the monthly rent paid for the house and the distance of the house from the tarmac road. (2 marks)
 - (ii) Interpret the co-efficient of determination obtained in (c) (i) above. (2 marks)

QUESTION FIVE (20 MARKS)

- a) Explain each of the following measures as used in statistics: (6 marks)
 - i. mean;
 - ii. median;
 - iii. Standard deviation.
- b) A job vacancy was advertised in which an aptitude test was conducted in three subject areas to evaluate the applicants. The subjects are Mathematics, English and Social Studies. From past experience, it is known that 20% pass Mathematics, 60% pass English and 50% pass Social Studies. To pass the interview, an applicant must pass:
 - All the three subjects to qualify for management position:
 - Mathematics and another one subject to qualify for middle level position
 - Any one single subject to qualify for support staff position
 - i. Present this information in a tree diagram. (4 marks)
 - ii. A total of 250 applicants did the interview. Assuming that the applicants were randomly shortlisted for the interview, estimate the number of applicants who qualified for
 - For each position
 - The total number applicants who were qualified for positions. (10 marks)