



MACHAKOS UNIVERSITY

University Examinations 2018/2019

SCHOOL OF PURE AND APPLIED SCIENCES

DEPARTMENT OF MATHEMATICS, STATISTICS AND ACTUARIAL SCIENCE

FIRST YEAR SECOND SEMESTER EXAMINATION

BACHELOR OF SCIENCE IN ELECTRICAL AND ELECTRONICS ENGINEERING

BACHELOR OF SCIENCE IN MATHEMATICS AND COMPUTER SCIENCE

BACHELOR OF SCIENCE IN STATISTICS AND PROGRAMMING

BACHELOR OF ECONOMICS AND STATISTICS

BACHELOR OF SCIENCE (MATHEMATICS)

BACHELOR OF EDUCATION (SCIENCE)

BACHELOR OF EDUCATION (ARTS)

BACHELOR OF ARTS

SMA 160: INTRODUCTION TO PROBABILITY AND STATISTICS

DATE: 2/5/2019

TIME: 8:30 – 10:30 AM

Instructions to the Candidate:

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

1. 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) Recent observations show that, 5% of the Machakos University male postgraduate students have prostate cancer. The university clinic laboratory can detect 89% of the cancer if it exists. However, it gives 12% false positive results where cancer does not exist. Suppose a male postgraduate student volunteers at random for a prostate cancer test in the university laboratory, determine the following probabilities:

- i) That the result was positive;
 - ii) That the student had prostate cancer given that the result was positive;
 - iii) That the student had no prostate cancer given the result was negative
 - iv) That the student was misclassified. (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 4 members is to be formed from a group of 6 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)
2. 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 any **three** of the measurement scales used in the classification of statistical data giving two examples in each case. (6 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 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)
3. 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:
- (i) The normal equations;
- (ii) 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) A researcher on land economics formulation two models to be used in estimating the price of a plot of land based on a specified set of input variables. In order to evaluate the two models, he collected data on the actual prices of a random sample of 12 plots and then obtained two estimates for each plot using each of the two models. The amount in ten thousand Kenya shillings was recorded for actual prices and also for the estimated prices from the two models as shown in the table below:

Plot	A	B	C	D	E	F	G	H	J	K	L	M
Actual	15	25	30	18	62	30	78	68	24	55	66	58
Model 1	18	23	32	25	34	64	82	90	45	56	60	48
Model 2	16	36	45	24	38	54	72	60	45	32	39	76

- i) Compute the Spearman's rank correlation co-efficient between the actual price and the price estimate by Model 1, and between the actual price and the price estimated by Model 2. (8 marks)
- ii) 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)

4. 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 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. (1 mark)
- 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.
- ii) Interpret the co-efficient of determination obtained in (c) (i) above. (3 marks)
5. a) Explain **three** properties of the linear least squares regression line. (5 marks)
- b) The data given below represents monthly salary for two samples of employees of a given bank based in Kenya and in South Africa. The salary for employees in Kenya is in Kenya shillings while those in South Africa have salary in Rands.

Country	Sample size	Mean	Std deviation
Kenya	150	75	10
South Africa	100	65	12

Determine each of the following measures about the monthly salaries:

- i) Country with higher dispersion in salary.
- ii) Combined standard deviation of the salaries. (6 marks)

- c) 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 any other 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. (3 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 each position. (6 marks)