KIRIRI WOMENS' UNIVERSITY OF SCIENCE AND TECHNOLOGY
UNIVERSITY EXAMINATIONS, 2024/2025 ACADEMIC YEAR
FIRST YEAR, FIRST SEMESTER EXAMINATION
FOR THE DEGREE OF BACHELOR OF SCIENCE(COMPUTER SCIENCE)

KMA 2102: INTRODUCTION TO PROBABILITY AND STATISTICS

Date: 6TH DECEMBER 2024
Time: 8:30AM-10:30 AM

INSTRUCTIONS TO CANDIDATES
ANSWER QUESTION ONE (COMPULSORY) AND ANY OTHER TWO QUESTIONS

QUESTION ONE: COMPULSORY (30 MARKS)

a) Distinguish between the following terms
   i. Primary data sources and secondary data sources **(2 Marks)**
   ii. Descriptive statistics and inferential statistics **(2 Marks)**

b) The age distribution of all players in the national teams in a certain country is shown below.

Age (Years)	20-24	25-29	30-34	35-39	40-44	45-49	50-54
No. of Players	11	24	30	18	11	5	1

Compute the coefficient of skewness and examine whether age distribution exhibits normal distribution. **(5 Marks)**

c) In a factory, machine A produces 55% of the output while machine B and machine C produces 25% and 20% of the output respectively. 2% of the output of machine A is defective while machine B and C has defective proportions of 1.5% and 1.8%. An item is drawn at random from the day's output and is found to be defective. Find the probability that it was produced by Machine B **(4 Marks)**

d) In studying the relationship between two factors X and Y, summary of the collected data was summarized as follows

$$\sum X = 725, \ \sum Y = 1011, \ \sum XY = 61685, \ \sum X^2 = 44475, \ \sum Y^2 = 85905, \ n = 12$$

   i. Find the equation of the regression line **(5 Marks)**
   ii. Compute the correlation coefficient and interpret the value obtained **(4 Marks)**

e) Suppose that lengths of components produced by a machine are as follows:
      43.8, 51.8, 53.4, 40.1, 36.8, 34.8, 37.5, 31.6, 44.7, 51.5, 45.3, 45.3, 46.0, 39.9 and 42.3.
Calculate:
   i. Harmonic mean of the observations **(4 Marks)**
   ii. Standard deviation using the harmonic mean obtained. **(4 Marks)**

QUESTION TWO: (20 MARKS)

a) Explain any three graphical methods that can be used to represent data **(3 Marks)**

b) In a statistics class, performance of 100 students in a CAT (out of 50) is shown in the table below

Score	0-4	5-9	10-14	15-19	20-24	25-29	30-34	35-39	40-44
Frequency	1	2	15	21	26	28	4	2	1

Use the data to obtain the following:
   i. Modal score **(3 Marks)**

ii. Semi-interquartile range **(6 Marks)**
iii. 2nd central moment **(6 Marks)**
iv. Coefficient of variation **(2 Marks)**

QUESTION THREE: (20 MARKS)

a) An internet provision company sells three different internet products. The company records the number of new weekly subscriptions as shown in the table below

Product	Week 1	Week 2	Week 3	Week 4	Week 5	Week 6	Week 7
Product I	19	20	21	24	25	21	20
Product II	15	16	18	19	16	16	14
Product III	27	28	25	31	32	25	28

For each product, calculate:
i. Mean using a suitable assumed value **(5 Marks)**
ii. Standard deviation using the actual means obtained in (i) above **(5 Marks)**
iii. Use coefficient of variation to determine which product has a stable number of subscriptions **(5 Marks)**

b) In studying the association between two variables X and Y, a researcher obtained the following data:
$$\sum X = 169, \ \sum Y = 130, \ \sum XY = 3006, \ \sum X^2 = 3856, \ \sum Y^2 = 2370, \ n = 10$$
From the information, compute the coefficient of determination and comment on your result **(5 Marks)**

QUESTION FOUR: (20 MARKS)

Consider the following data showing bank account balances (in '000) as reported by a certain audit firm

113	106	100	91	100	96	112	107	125	98
105	101	121	101	104	92	101	98	97	110
103	127	94	120	101	93	98	93	93	109
100	91	108	111	95	103	95	118	106	123
102	98	117	98	117	104	98	102	93	105

a) Construct a frequency distribution table with classes 90-94, 95-99, 100-104 etc **(3 Marks)**
b) Using the frequency distribution table obtained above, obtain
i. Geometric mean **(4 Marks)**
ii. Ogive curve and estimate the median value from the curve **(5 Marks)**
iii. Calculate median using the median formula and compare this value and the estimated median value using the curve **(5 Marks)**
c) A toy is rejected if the design is faulty or not. The probability that the design is faulty is 0.1 and that the toy is rejected if the design is faulty is 0.95 and otherwise 0.45. If a toy is rejected, what is the probability that it is due to faulty design? **(3 Marks)**

QUESTION FIVE: (20 MARKS)

a) Given the data below of the

Classes	21-25	26-30	31-35	36-40	41-45	46-50	51-55	56-60	61-65	66-70
Frequency	7	13	9	18	12	21	11	9	7	3

Using the data provided, obtain:
i. Mean absolute deviation **(5 Marks)**
ii. Histogram and comment on the distribution **(4 Marks)**
iii. Coefficient of skewness and comment on the value obtained. Compare whether a similar conclusion is made in (ii) and (iii) **(5 Marks)**

b) Use the data provided in the table below to estimate the regression line. Clearly show all your workings. **(6 Marks)**

X	42	18	22	30	60	56
Y	45	33	10	28	45	60