LAIKIPIA UNIVERSITY

UNIVERSITY EXAMINATIONS

SECOND SEMESTER 2023/2024 ACADEMIC YEAR

FIRST YEAR EXAMINATION FOR THE DEGREE OF BACHELOR OF EDUCATION (SCIENCE/ARTS)

STAT 111: INTRODUCTION TO PROBABILITY AND STATISTICS I

STREAM: R ***TIME: 2 HRS***

DAY: TUESDAY [11.30A.M -1.30P.M] ***DATE: 16/04/2024***

THIS QUESTION PAPER CONSISTS OF FIVE (5) PAGES

PLEASE DO NOT OPEN UNTIL THE INVIGILATOR SAYS SO.

INSTRUCTIONS: **ANSWER QUESTION ONE AND ANY OTHER TWO**

QUESTION ONE [30 MARKS]

a) Explain the following terms **[10 Marks]**

i. Convenience sampling
ii. Multistage sampling
iii. Sampling frame
iv. Mutually exclusive events
v. Random experiment

b) The table below gives the number of thunderstorms reported in a particular summer month by 100 meteorological stations.

i.

Number of thunderstorms	0	1	2	3	4	5
Number of stations:	22	37	20	13	6	2

i) Calculate the sample mean number of thunderstorms. **[2 Marks]**

ii) Calculate the sample median number of thunderstorms. **[2 Marks]**

iii) Comment briefly on the comparison of the mean and the median. **[2 Marks]**

c) A random variable X has the following probability distribution

X	0	1	2	3	4	5	6	7
P(x)	0	K	2k	2k	3k	K^2	$2k^2$	$7k^2+k$

d) Determine;

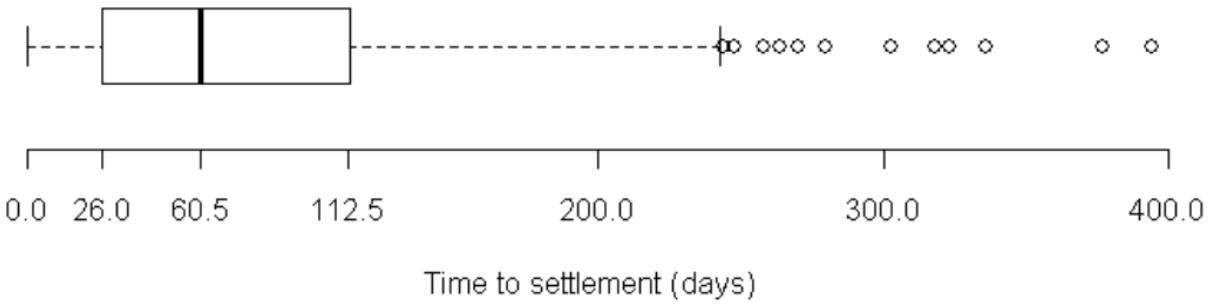
a) The value of constant **k** **[2 Marks]**

b) The expectation of X **[3 Marks]**

c) Standard deviation of X **[3 Marks]**

e) Explain the disadvantage of using dispersion as a measure of variation **[2 Marks]**

f) Data were collected on the time (in days) until each of 200 claims is settled by the insurer in a certain insurance portfolio. A boxplot of the data is shown below.



Calculate the median and the interquartile range of the data using the plot. **[2 Marks]**

d) An average weight of 10 students was calculated to be 65. Later it was discovered that one weight was misread as 40kg instead of 80kg.
Calculate the correct average weight. **[2 Marks]**

QUESTION TWO [20 MARKS]

a) A sample of 20 claim amounts (£) on a group of household policies gave the following data summaries:
$\sum_{i=1}^{n} x_i = 3256$ and $\sum_{i=i}^{n} x_i^{\,2}$=866600

i) Calculate the sample mean and standard deviation for these claim amounts. **[2 Marks]**

ii) Comment on the skewness of the distribution of these claim amounts, giving reasons for your answer. **[2 Marks]**

b) Aster and Almaz were asked to rank 7 different types of lipsticks, check if there is a correlation between the tests of the two ladies. **[6 Marks]**

Lipstick		A	B	C	D	E	F	G
Aster		2	1	4	3	5	7	6
Almaz		1	3	2	4	5	6	7

c) Construct a stem and leaf plot for the 20 test scores given below **[3 Marks]**

78, 91, 74, 74, 82, 82, 66, 75, 94, 96, 71, 78, 64, 84, 88, 79, 55, 71, 80, 83

d) In a tank containing 10 fishes, there are three yellow and seven black fishes. We select three fishes at random.

i). What is the probability that exactly one yellow fish gets selected? **[2 Marks]**

ii). What is the probability that at most one yellow fish gets selected? **[3 Marks]**

iii). What is the probability that at least one yellow fish gets selected? **[2 Marks]**

QUESTION THREE [20 MARKS]

a) A survey, carried out at a major flower and gardening show, was concerned with the association between the intention to return to the show next year and the purchase of goods at this year`s show. There were 220 people interviewed and of these 101 had made a purchase; 69 of these people said they intended to return next year. Of the 119 who had not made a purchase, 68 said they intended to return next year. Suppose one of the 220 people surveyed is selected at random. Calculate the probabilities that the selected person:

i) Intends to return next year, given that he/she has made a purchase **[2 Marks]**
ii) Intends to return next year, given that he/she has not made a purchase **[2 Marks]**
iii) Has made a purchase, given that he/she intends to return next year [**2 Marks]**

b) State four disadvantages of interviewing as a method of data collection **[4 Marks]**

c) Find the harmonic and geometric mean for the data in the frequency table below; **[6 Marks]**

x	10	12	13	15	16
f	2	4	10	6	4

d) Given n=10, $\sigma_x = 5.4$, $\sigma_y = 6.2$ and the sum of the deviation from the mean of X and Y is 66. Find the correlation coefficient **[4 Marks]**

QUESTION FOUR [20 MARKS]

a) Explain the following terms as used in statistics [**8 Marks]**

i). Regression analysis
ii). Correlation analysis
iii). Sampling
iv). Variable

b) A development engineer examined the relationship between the speed a vehicle is travelling (in miles per hour, mph), and the stopping distance (in metres, m) for a new braking system fitted to the vehicle. The following data were obtained in a series of independent tests conducted on a particular type of vehicle under identical road conditions.

c)

Speed of vehicle (*x*):	10	20	30	40	50	60	70
Stopping distance (y):	5	10	23	34	40	54	75

i). Construct a scatterplot of the data, and comment on whether a linear regression is appropriate to model the relationship between the stopping distance and speed. **[4 Marks]**

ii). Calculate the equation of the least-squares fitted regression line and interpret. **[8 Marks]**

QUESTION FIVE [20 MARKS]

a) Find the moment coefficient of skewness and kurtosis for the data in the frequency table below; **[6 Marks]**

X	10	12	14	18	20
F	3	5	10	5	6

b) Consider three events *A*, *B*, and *C* for which *A* and *C* are independent, and *B* and *C* are mutually exclusive. You are given the probabilities $P(A) = 0.3$, $P(B) = 0.5$, $P(C) = 0.2$ and $P(A \cap B) = 0.1$. Find the probability that none of *A*, *B*, or *C* occurs. **[3 Marks]**

c) The random variable *X* has probability density function

$$f(x) = k(1-x)(1+x),\ 0 < x < 1$$

Where, *k* is a positive constant,

i) Show that $k = 1.5$. **[2 Marks]**

ii) Calculate the probability $P(X > 0.25)$. **[2 Marks]**

d) An analysis of the monthly wages paid to workers in two firms A and B belonging to the same industry gives the following results;

Value	Firm A	Firm B
Mean wage	52.5	47.5
variance	100	121

In which firm A or B is there a greater variability in individual wages? **[5 Marks]**

e) Explain the difference between multiple and component bar charts **[2 Marks]**