(Knowledge for Development)

KIBABII UNIVERSITY

UNIVERSITY EXAMINATIONS

2016/2017 ACADEMIC YEAR

FIRST YEAR SECOND SEMESTER

SPECIAL/SUPPLEMENTARY EXAMINATION

FOR THE DEGREE OF BACHELOR OF SCIENCE

COURSE CODE: STA 114

COURSE TITLE: INTRODUCTION TO STATISTICS

DATE: 21/09/17 **TIME:** 8 AM -10 AM

INSTRUCTIONS TO CANDIDATES

Answer Question One and Any other TWO Questions

TIME: 2 Hours

This Paper Consists of 5 Printed Pages. Please Turn Over.

QUESTION ONE COMPULSORY (30 marks)

a. What do you understand by the following statistics terms:
   I. Statistics
   II. Population
   III. Index number
   IV. Nominaldata (4marks)

b. State any two methods used in collecting primary data. (2marks)

c. What can be said about a sample of observation whose standard deviation is zero? (1 mark)

d. The ages of individuals in a manufacturing company is given as below:

62,21,26,32,56,36,37,39,40,54,42,44,61,68,33,56,57,37,52,39,40,54,43,43,63,30,34,68,35,38,50,38,52 ,41,51,44,41,42,43,45,46,45,47,48,49,45,46,48.

   I. Tabulate a grouped frequency distribution for the above data. (2marks)

   II. Construct a frequency curve for the above grouped frequency distribution and comment on the nature of the distribution. (3marks)

   III. Construct an ogive curve and use it to get the inter-quartile range. (4marks)

e. The heights (cm) and weights (kgs) of a sample of eight students are given below.

heights(x)	177	163	168	174	184	184	171	173
weights (y)	71	67	62	73	80	85	69	77

Find the Pearson's Correlation Coefficient between heights and weights. What conclusion can you make? (5marks)

f. For the data given below compute Laspeyer's Paasche's and Fischer's ideal indices and Bowleys method.

Commodity	Price 1975	Quantity 1975	Price 1980	Quantity 1980
A	**4**	**20**	**6**	**10**
B	**3**	**15**	**5**	**20**
C	**2**	**25**	**3**	**15**
D	**5**	**10**	**4**	**40**

(9marks)

QUESTION 2 (20 MARKS)

a. Differentiate between the following terms as used in statistics:
   I. Descriptive statistics and inferential statistics (2marks)
   II. Skewness and Kurtosis (2marks)
   III. Population and sample (2marks)

b. Give any two limitations of arithmetic mean as a measure of central tendency (2marks)

c. The following data shows how fertilizer used affect the sugarcane yield for equal size plots:

X-Fertilizer	0.3	0.6	0.9	1.2	1.5	1.8	2.1	2.4
y-Sugar Yield	10	15	30	35	25	30	50	45

Plot a scatter diagram for the dependent variable against the independent variable and comment on the nature of the relationship that exist (3marks)

d. Describe under what circumstances is harmonic mean preferred as a measure of central tendency? (2marks)

e. The blood pressure levels of first year male education students (in mmHg) in a particular university is given below

Class limit	Frequency
90-99	2
100-109	6
110-119	17
120-129	16
130-139	12
140-149	6
150-159	1

Calculate

i. Geometric and harmonic mean (4 marks)

ii. The mode (3 Marks)

QUESTION 3 (20 MARKS)

a. Show that variance is independent of change of origin (4 Marks)

b. Areas of material surfaces with rusts from a sample of 15 houses are as follows: 101,105,110,114,115,124,125,125,130,133,135,136,137,140,145. Find the standard deviation of the distribution and comment on it. (4 Marks)

c. An analysis of monthly salary paid to the workers in two firms A and B belonging to the same industry gives the following results.

	Firm A	Firm B
Number of workers	500	600
Average monthly salary in $	186	175
Variance of distribution salaries	81	100

i. In which firm A or B has larger salary bill? (2 Marks)

ii. In which firm A or B, is there greater variability in individual salaries?(2 Marks)

iii. Calculate the combined average monthly wage for all the workers in the firms A and B (3 Marks)

d. The highway police patrol set up radar check point and recorded the speed in miles per hour of a random sample of 50 cars that passed the check points in one hour. The observations were as follows:-

77	66	65	55	48
56	50	65	75	67
76	58	50	65	60
65	60	51	68	76
68	77	63	65	52
52	63	65	80	70
65	81	70	63	53
45	65	55	71	64
55	70	64	45	66
64	40	66	55	71

Construct stem and leaf distribution for the data and use it to establish the quartiles. (5 Marks)

QUESTION 4 (20 MARKS)

a. Let $x_1, x_2 \ldots x_n$ be a sample data of size n whose mean is $\bar{x}$ and variance s^2. Show that

$$s^2 = \frac{\sum(x_1 - \bar{x})^2}{n-1} \text{ can be written as } s^2 = \frac{n\sum x_i^2 - \left(\sum x_i\right)^2}{n(n-1)}$$ **(3 marks)**

b. Compute the value of the correlation coefficient for the data obtained in the study of the number of absences and the final grade of the seven students in the statistics class as shown: (4marks)

Student	No.of absencesX	Final Grade Y	xy	X2	Y2
A	6	82	492	36	6724
B	2	86	172	4	7396
C	15	43	645	225	1849
D	9	74	666	81	5476
E	12	58	696	144	3364
F	5	90	450	25	8100
G	8	78	624	64	6084

c. Using the following frequency distribute table

Marks	**No. of students**
0-10	2
10-20	10
20-30	16
30-40	24
40-50	33
50-60	16
60-70	12
70-80	10
80-90	6
90-100	1

Find the

i. Arithmetic mean (2 marks)
ii. Median (2 marks)
iii. 40^{th} percentile (2 marks)
iv. Coefficient of skewness and kurtosis based on method of moments , hence comment on the skewness of the distribution (7 marks)

QUESTION 5 (20 MARKS)

a. Show that the variance is the minimum value of mean square deviation. (4 marks)
b. The director of admissions of a small college admission a newly designed entrance test to 20 students selected at random from the freshman class in a study to determine whether a student's grade point average (GPA) at the end of the freshman year (Y) can be predicted from the entrance test score (X). The summary of the results are as given below:-
Assuming a simple linear regression model best fits the data values, obtain the least squares estimates for this model, and hence write down the model relating Y and X. interpret it. (5 marks)
c. State three properties of testing the adequacy of a good index number formula (3 marks)

d. The table below shows the weighted aggregate price index number for data between 1980 to 1981 as base year

Commodity	Price		Quantity	
	1980	1981	1980	1981
A	10	12	20	22
B	8	8	16	18
C	5	6	10	11
D	4	4	7	8

Computer the Fisher's ideal index number (3mks)

e. The following marks belong to 99 students of a municipality school XYZ for the visually impaired.

marks	**Number of students**
0-10	10
10-20	?
20-30	25
30-40	30
40-50	?
50-60	10

On later analysis it discovered that two class intervals frequencies were missing. If the median was found to be 30, find the missing frequencies and the hence the model marks of the students. (5 marks)