MASENO UNIVERSITY

UNIVERSITY EXAMINATIONS 2017/2018

FIRST YEAR FIRST SEMESTER EXAMINATION FOR THE DEGREE OF BACHELOR OF SCIENCE IN APPLIED STATISTICS, ACTUARIAL SCIENCE, MATHEMTICAL SCIENCE AND MATHEMATICS AND ECONOMICS WITH INFORMATION TECHNOLOGY

MAIN CAMPUS

MMA 104/MAS 101: DESCRIPTIVE STATISTICS

Date: 21st February, 2018 Time: 12.00 - 3.00pm

INSTRUCTIONS:

- Answer Question ONE and any other TWO.

QUESTION 1 (30 Marks) - COMPULSORY

a) An adult evening class has 14 students. The ages of these students have a mean of 31.2 years and a standard deviation of 7.4 years. A new student who is exactly 42 years old joins the class. Calculate the mean and standard deviation of the 15 students now in the group. **(7 Marks)**

b) A tennis coach believes that taller players are generally capable of hitting faster serves. To investigate this hypothesis, he collects data on the 20 adult male players he coaches. The height, h, in metres and the speed of each player's fastest serve, v, in miles per hour were recorded and summarized as follows:

$$\sum h = 36.22, \sum v = 2275, \sum h^2 = 65.7396, \sum v^2 = 259853, \sum hv = 4128.03$$

i) Calculate the product moment correlation coefficient for these data. **(5 marks)**

ii) Comment on the coach's hypothesis. **(2 marks)**

c) For a project, a student asked 40 people to draw two straight lines with what they thought was an angle of 75° between them, using just a ruler and a pencil. She then measured the size of the angles that had been drawn and her data are summarized in this stem and leaf diagram.

Angle	(6|4 means 64°)	Totals
4	1	(1)
4		(0)
5	0 2 4	(3)
5	5 8 9	(3)
6	1 1 3 3 4	(5)
6	5 5 7 8 9	(5)
7	0 1 1 2 3 3 4 4 4	(9)
7	5 6 6 7 7 9 9	(7)
8	0 1 1 3 4	(5)
8	5 6	(2)

i) Find the median and quartiles of these data. **(4 marks)**

ii) Use the 1.5IQR criterion to determine if there are any outliers in these data **(4 marks)**

iii) Draw a box plot representing these data on graph paper, **(5 marks)**

iv) Describe the skewness of the distribution and suggest a reason for it. **(3 marks)**

QUESTION 2 (20 Marks)

a) The three measures of central tendency are the mode, median and mean. State when each measure of tendency would be used to describe data **(9Marks)**

b) A soccer fan collected data on the number of minutes of league football, m, played by each team in the four main divisions before first scoring a goal at the start of a new season. Her results are shown in the table below.

m (minutes)	Number of teams
$0 \leq m < 40$	36
$40 \leq m < 80$	28
$80 \leq m < 120$	10
$120 \leq m < 160$	4
$160 \leq m < 200$	5
$200 \leq m < 300$	4
$300 \leq m < 400$	2
$400 \leq m < 600$	3

i) Calculate estimates of the mean and standard deviation of these data. **(8 marks)**

ii) Explain why the mean and standard deviation might not be the best summary statistics to use with these data. **(2 marks)**

iii) Suggest alternative summary statistics that would better represent these data. **(1 mark)**

QUESTION 3 (20 Marks)

The number of patients attending a hospital trauma clinic each day was recorded over several months, giving the data in the table below.

Number of patients	10 - 19	20 - 29	30 - 34	35 - 39	40 - 44	45 - 49	50 - 69
Frequency	2	18	24	30	27	14	5

a) These data are represented by a histogram. Given that the bar representing the 20 - 29 group is 2 cm wide and 7.2 cm high, calculate the dimensions of the bars representing the groups

i) 30 - 34

ii) 50 - 69

(6 marks)

b) Use linear interpolation to estimate the median and quartiles of these data.

(9 marks)

c) The lowest and highest numbers of patients recorded were 14 and 67 respectively. Represent these data with a boxplot drawn on graph paper and describe the skewness of the distribution. **(5 marks)**

QUESTION 4 (20 Marks)

a) Define a categorical variable **(2 Marks)**

b) Distinguish between the nominal and ordinal level of measurement **(2 Marks)**

c) Give two examples each of variables measured in nominal and ordinal scale

(4 Marks)

d) A company started and managed by business students is selling campus calendars. The students have conducted a market survey with the various campus constituents to determine sales potential and identify which market segments should be targeted. (Should they advertise in the alumni magazine and/or the local newspaper?) The following table shows the results of the market survey.

		Buying Likelihood			
		Unlikely	**Moderately Likely**	**Very Likely**	**Total**
Campus Group	Students	197	388	320	**905**
	Faculty/Staff	103	137	98	**338**
	Alumni	20	18	18	**56**
	Town Residents	13	58	45	**116**
	Total	**333**	**601**	**481**	**1415**

i) Represent the above information using an appropriate graph **(4 Marks)**

ii) What percent of all these respondents are alumni? **(1 Marks)**

iii) What percent of these respondents are very likely to buy the calendar?

(1 Marks)

iv) What percent of the respondents who are very likely to buy the calendar are alumni? **(1 Marks)**

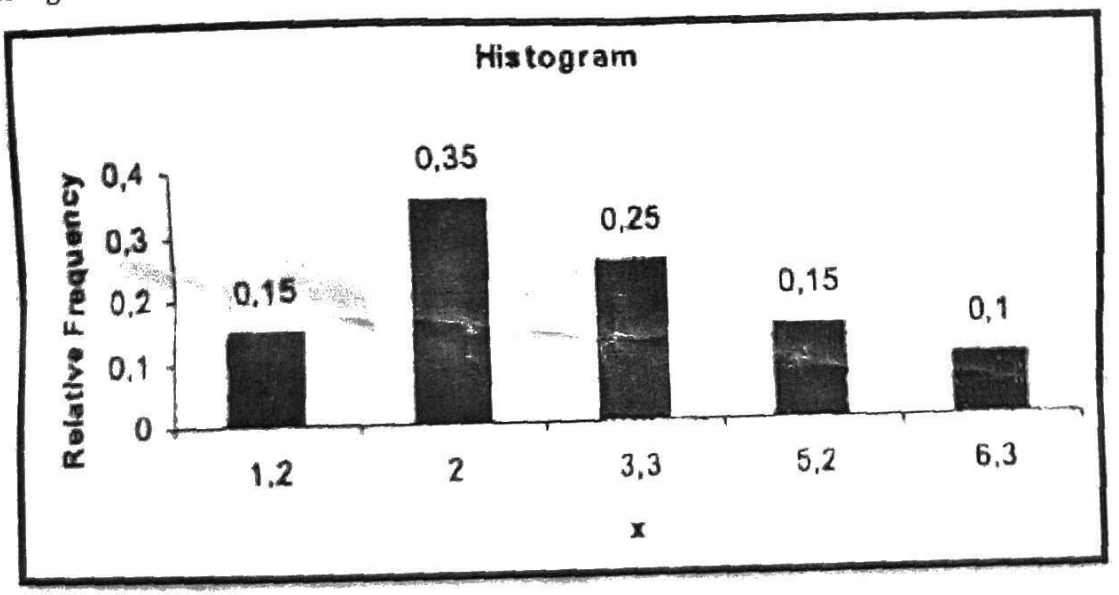
v) Of the alumni, what percent are very likely to buy the calendar? **(1 Marks)**

vi) What is the marginal distribution of the campus constituents? **(2 Marks)**

vii) What is the conditional distribution of the campus constituents among those very likely to buy the calendar? **(2 Marks)**

QUESTION 5 (20 Marks)

The histogram of a discrete data set consisting of 100 measurements is given below:



Determine

a) the mode, median and mean (5 Marks)

b) the inter-quartile range (5 Marks)

c) the frequency of 2 (2 Marks)

d) the sum of the data points (3 Marks)

e) the value of $\sum_{i=1}^{N} x_i^2$ given that the variance is 2.687 (5 Marks)