



**MASENO UNIVERSITY**  
**UNIVERSITY EXAMINATIONS 2016/2017**

**FIRST YEAR FIRST SEMESTER EXAMINATIONS FOR THE  
DEGREE OF BACHELOR OF EDUCATION WITH INFORMATION  
TECHNOLOGY**

**CITY CAMPUS**

**CIT 102: PROBABILITY AND STATISTICS**

Date: 15<sup>th</sup> June, 2017


Time: 5.30 - 8.30 pm

---

**INSTRUCTIONS:**

- Answer question ONE and any other TWO questions.

MASENO UNIVERSITY

ISO 9001:2008 CERTIFIED 

Instruction: Answer Question One and any other two questions

**Question 1 (30 Marks)**

a) Data can be classified into four levels of measurement.

i) State and give a brief description of these four levels of measurement

---

(4 Marks)

ii) Indicate which level of measurement is being used in the given scenario:

a. The teacher of a class of third graders records the eye colour of each student.

b. The teacher of a class of third graders records the letter grade for mathematics for each student.

---

c. A meteorologist compiles a list of temperatures in degrees Celsius for the month of May

---

d. A meteorologist compiles a list of temperatures in degrees Kelvin for the month of May

(4 Marks)

iii) Explain why it is necessary for a statistician to verify the level of data before carrying out any analysis.

(2 Marks)

b) The students in a class were each asked to write down how many CDs they owned. The student with the least number of CDs had 14 and all but one of the others owned 60 or fewer. The remaining student owned 65. The quartiles for the class were 30, 34 and 42 respectively. Outliers are defined to be any values outside the limits of  $1.5(Q_3 - Q_1)$  below the lower

---

quartile or above the upper quartile. On graph paper draw a box plot to represent these data, indicating clearly any outliers. (7 Marks)

c) A computer can generate random numbers which are either 0 or 2. On a particular occasion, it generates a set of numbers which consists of 23 zeros and 17 twos. Find the mean and variance of this set of 40 numbers.

(4 Marks)

d) The following data represent the daily ticket sales at a small theatre during three weeks.

52, 73, 34, 85, 62, 79, 89, 50, 45, 83, 84, 91, 85, 84, 87, 44, 86, 41, 35, 73, 86.

i) Construct a stem-and-leaf diagram to illustrate the data. (3 Marks)

ii) Use your diagram to find the median of the data. (1 Mark)

e) An experiment carried out by a student yielded pairs of  $(x, y)$  observations such that

$$\bar{x} = 36, \bar{y} = 28.6, S_{xx} = 4402, S_{xy} = 3477.6$$

i) Determine the equation of the regression line of  $y$  on  $x$  in the form  $y = a + bx$ . Give your values of  $a$  and  $b$  to 2 decimal places

(3 Marks)

ii) Find the value of  $y$  when  $x = 45$

(2 Marks)

### Question 2 (20 Marks)

a) Explain the meaning of the following terms as used in statistics

i) a sample space (1 Marks)

ii) an event (1 Marks)

b) Two events  $A$  and  $B$  are independent, such that  $p(A) = \frac{1}{3}$  and  $p(B) = \frac{1}{4}$ .

Find

i)  $p(A \cap B)$  (2 Marks)

ii)  $p(A|B)$  (2 Marks)

iii)  $p(A \cup B)$  (2 Marks)

c) A discrete random variable  $X$  takes only positive integer values. It has a cumulative distribution function  $F(x) = P(X \leq x)$  defined in the table below.

---

$X$	1	2	3	4	5	6	7	8
$F(x)$	0.1	0.2	0.25	0.4	0.5	0.6	0.75	1

- i) Determine the probability function,  $p(X = x)$ , of  $X$ . (3 Marks)  
ii) Calculate  $E(X)$  and show that  $Var(X) = 5.76$  (6 Marks)  
iii) Given that  $Y = 2X + 3$ , find the mean and variance of  $Y$ . (3 Marks)
- 

**Question Three (20 Marks)**

- a) Define the following terms as used in statistics
- i) Population
  - ii) Sample
  - iii) Variable
  - iv) Parameter
  - v) Statistic (10 Marks)
- b) An experiment was conducted to test the effect of a new drug on a viral infection. After the infection was induced in 100 mice, the mice were randomly split into two groups of 50. The first group, the control group, received no treatment for the infection, and the second group received the drug. After a 30-day period, the proportion of survivors,  $\hat{p}_1$  and  $\hat{p}_2$ , in the two groups was found to be 0.36 and 0.60 respectively.
- i) Is there sufficient evidence to indicate that the drug is effective in treating the infection? Use  $\alpha = 0.05$  (7 Marks)
- ii) Use a 95% confidence interval to estimate the actual difference in the survival rates for the treated versus the control groups
- 

(3 Marks)

**Question Four (20 Marks)**

- a) Define the following terms as used in statistics
- i) Confidence Interval
  - ii) Hypothesis
  - iii) Level of significance
  - iv) Type I error
  - v) Type II error (10 Marks)
- b) The mean lifetime of 100 components in a sample is 1,570 hours and their standard deviation is 120 hours.
- i) Give a 95% confidence interval of the population mean (5 Marks)
  - ii) Is it likely the sample comes from a population whose mean is 1,600 hours? Test at 1% level of significance (5 Marks)

**Question Five (20 Marks)**

In order to study the relationship between exam preparation time,  $x$ , and the score received on the final exam,  $y$ , six students were asked to keep a record of their exam preparation time (in hours). The results are as follows:

Preparation time $x$	1.7	2.1	3.2	3.0	4.0	3.5
Score on exam $y$	65	63	71	69	75	72

Note:  $\sum x = 17.5$ ,  $\sum y = 415$ ,  $\sum x^2 = 54.79$ ,  $\sum y^2 = 28805$ ,  $\sum xy = 1229$

- a) Calculate the correlation coefficient,  $r$ . (6 Marks)
- b) Interpret the correlation coefficient,  $r$  calculated in (a) (2 Marks)
- c) Find the line-of-best-fit (linear regression line). (8 Marks)
- d) Estimate a student's score on the exam if they studied 2.5 hours. (4 Marks)