

FORM TP 2016139



TEST CODE 02105020

MAY/JUNE 2016

CARIBBEAN EXAMINATIONS COUNCIL

CARIBBEAN ADVANCED PROFICIENCY EXAMINATION®

APPLIED MATHEMATICS

STATISTICAL ANALYSIS

UNIT 1 – Paper 02

2 hours 30 minutes

READ THE FOLLOWING INSTRUCTIONS CAREFULLY.

1. This paper consists of THREE sections. Answer ALL questions from the THREE sections.
2. Write your answers in the spaces provided in this booklet.
3. Do NOT write in the margins.
4. Unless otherwise stated in the question, all numerical answers MUST be given exactly OR to three significant figures as appropriate.
5. If you need to rewrite any answer and there is not enough space to do so on the original page, you must use the extra page(s) provided at the back of this booklet. **Remember to draw a line through your original answer.**
6. **If you use the extra page(s) you MUST write the question number clearly in the box provided at the top of the extra page(s) and, where relevant, include the question part beside the answer.**

Examination Materials:

Mathematical formulae and tables (Revised 2010)

Electronic calculator

Ruler and graph paper

DO NOT TURN THIS PAGE UNTIL YOU ARE TOLD TO DO SO.

Copyright © 2014 Caribbean Examinations Council

All rights reserved.

“*”Barcode Area”*

Sequential Bar Code

SECTION A

MODULE 1: COLLECTING AND DESCRIBING DATA

Answer BOTH questions.

1. (a) A random sample of 25 students was taken from the fifth forms of a certain school in proportion to the number of students in each form. The data collected were classified as the following three variables:

X: the number of CSEC subjects each student was studying

Y: the favourite subject being studied

Z: the distance each student travelled from home to school each day

State

- (i) the population of interest

..... [1 mark]

- (ii) the sampling method that was used to collect the data

..... [1 mark]

- (iii) the variable which is qualitative

..... [1 mark]

- (iv) the variable which is continuous

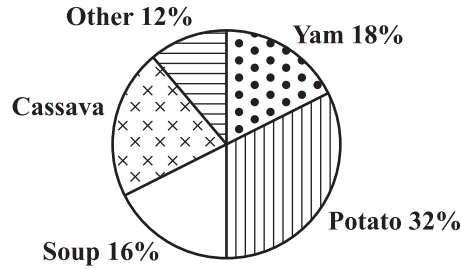
..... [1 mark]

- (v) the variable which is discrete.

..... [1 mark]

GO ON TO THE NEXT PAGE

- (b) The following pie chart, **not drawn to scale**, shows the distribution of some of the favourite foods of the residents at a home for the elderly. (Give ALL answers to the nearest whole number.)



- (i) Calculate the angle of the sector which represents those persons who love potato.

.....
.....

[2 marks]

- (ii) If 10 persons from the home love yam, determine the total number of persons who live in the home.

.....
.....
.....

[3 marks]

- (iii) Calculate the number of persons who love cassava.

.....
.....
.....

[3 marks]

GO ON TO THE NEXT PAGE

- (c) The following table shows the frequency distribution of the ages of persons who were booked to go on an island tour.

Age of Persons (years)	Number of Persons
5–14	7
15–24	16
25–34	28
35–44	18
45–54	13
55–64	10
65–84	8

- (i) State a disadvantage of displaying this data in groups as shown.

.....
.....
.....
.....

[1 mark]

- (ii) **On the grid provided on page 7**, draw a clearly labelled histogram showing the information in the table above.

Use a scale of 1 cm \equiv 5 years and 2 cm \equiv 5 persons.

[8 marks]

GO ON TO THE NEXT PAGE

(iii) Use the histogram to estimate the mode of the ages.

.....

.....

.....

.....

[3 marks]

Total 25 marks

DO NOT WRITE IN THIS AREA

2. (a) State which ONE of the mean, median and mode

(i) may be determined from qualitative data

..... [1 mark]

(ii) is affected by extreme values

..... [1 mark]

(iii) may have the largest value for a given set of data.

..... [1 mark]

(b) State which ONE of the range, interquartile range and variance

(i) is influenced most by extreme values

..... [1 mark]

(ii) may be obtained by squaring the standard deviation

..... [1 mark]

(iii) is the middle 50% of the distribution.

..... [1 mark]

DO NOT WRITE IN THIS AREA

GO ON TO THE NEXT PAGE

(c) A sample of 36 data values, x , gave $\sum(x - 45) = -148$ and $\sum(x - 45)^2 = 3\,089$.

(i) Calculate estimates for the mean.

.....
.....
.....
.....
.....

[3 marks]

(ii) Calculate the estimates for the standard deviation of the data.

.....
.....
.....
.....

[3 marks]

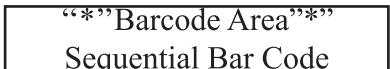
(iii) Another data value, $x = 29$, was added to the values. Calculate values for the mean and the standard deviation of the 37 values.

.....
.....
.....
.....
.....
.....

[5 marks]

DO NOT WRITE IN THIS AREA

GO ON TO THE NEXT PAGE



(d) The frequency distribution of the number of marbles per packet in a random sample of 50 packets is shown in the following table.

Number of Marbles per Packet (x)	13	14	15	16	17
Frequency, (f)	2	9	19	12	8

(i) Determine the median number of marbles per packet.

.....
.....
[2 marks]

(ii) Determine the mode of the distribution.

.....
[1 mark]

(iii) Calculate the mean number of marbles per packet.

.....
.....
.....
.....
.....
.....
.....
.....
.....
[4 marks]

(iv) Describe the shape of the distribution.

.....
[1 mark]

Total 25 marks

GO ON TO THE NEXT PAGE

DO NOT WRITE IN THIS AREA

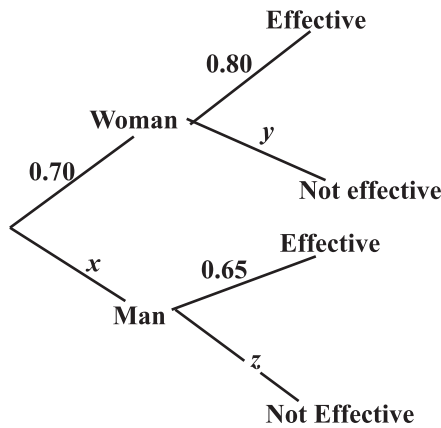
SECTION B

MODULE 2: MANAGING UNCERTAINTY

Answer BOTH questions.

3. (a) The effect of taking an antidepressant drug varies from person to person. It is known that 70% of the persons who use the drug are women. It is also known that 80% of the women and 65% of the men who use the drug find it to be effective.

The tree diagram below shows this information. The values not given are represented by x , y and z .



- (i) Calculate the values of x , y and z .

.....
.....
.....

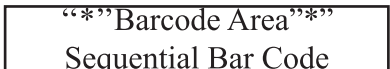
[3 marks]

- (ii) Calculate the probability that the drug is effective.

.....
.....
.....

[3 marks]

GO ON TO THE NEXT PAGE



DO NOT WRITE IN THIS AREA

- (iii) Given that the drug was effective, what is the probability that a person using the drug is a man?

.....

.....

.....

.....

[3 marks]

- (b) A department store is analysing the method of payment that customers use to determine whether there is a relation between the amount of money spent and the method of payment. The information is displayed in the following table.

	Cash	Debit Card	Credit Card
Under \$50	9	3	4
\$50–\$200	5	21	18
Over \$200	3	23	14

Calculate the probability that a person selected at random

- (i) pays for his/her item with cash

.....

.....

.....

.....

[2 marks]

DO NOT WRITE IN THIS AREA

(ii) makes a purchase that costs between \$50 and \$200

.....
.....
.....
.....

[2 marks]

(iii) makes a purchase over \$200 and pays with a debit card

.....
.....
.....
.....

[2 marks]

(iv) uses a credit card, given that the purchase was under \$50.

.....
.....
.....
.....

[2 marks]

DO NOT WRITE IN THIS AREA

GO ON TO THE NEXT PAGE

(c) Given two events R and S such that $P(R) = 0.6$, $P(S) = 0.5$ and $P(R \cap S) = 0.20$,

(i) calculate $P(R \cup S)$

.....
.....
.....
.....

[2 marks]

(ii) determine $P(R|S)$

.....
.....
.....
.....

[2 marks]

(iii) state, with reason, whether R and S are mutually exclusive

.....
.....
.....
.....

[2 marks]

(iv) state, with reason, whether R and S are independent.

.....
.....
.....
.....

[2 marks]

Total 25 marks

GO ON TO THE NEXT PAGE

DO NOT WRITE IN THIS AREA

4. (a) For EACH of the following, state, with reason, whether or not, it is a valid probability distribution.

(i)

x	0	1	2	3
$P(X = x)$	0.3	0.4	-0.2	0.5

.....
.....
[2 marks]

(ii)

x	0	1	2	3
$P(X = x)$	0.2	0.3	0.4	0.1

.....
.....
[2 marks]

(iii)

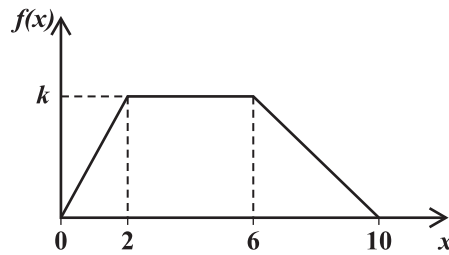
x	0	1	2	3
$P(X = x)$	0.25	0.46	0.24	0.18

.....
.....
[2 marks]

DO NOT WRITE IN THIS AREA

GO ON TO THE NEXT PAGE

- (b) The following graph, **not drawn to scale**, shows the probability density function, $f(x)$, of a continuous random variable, X .



- (i) Show that $k = \frac{1}{7}$.

.....

[2 marks]

- (ii) Determine $P(4 \leq X \leq 6)$.

.....

[3 marks]

- (c) A discrete random variable, X , has the following probability distribution.

x	1	2	4	8	16
$P(X = x)$	0.2	0.1	0.3	0.3	0.1

- (i) Calculate $\text{Var}(X)$.

.....

[6 marks]

GO ON TO THE NEXT PAGE

DO NOT WRITE IN THIS AREA

(ii) Construct a cumulative distribution table for the random variable X .

[2 marks]

(iii) Evaluate $P(X \geq 4)$.

.....
.....
.....

[1 mark]

DO NOT WRITE IN THIS AREA

GO ON TO THE NEXT PAGE

(d) Market research has shown that 60% of persons who are introduced to a certain product actually buy the product.

(i) If 80 persons are introduced to the product, determine the number of persons who are expected to buy the product.

.....
.....
.....
.....

[2 marks]

(ii) From a random sample of 15 persons who were introduced to the product, calculate the probability that **exactly** 9 will buy the product.

.....
.....
.....
.....
.....
.....

[3 marks]

Total 25 marks

DO NOT WRITE IN THIS AREA

SECTION C

MODULE 3: ANALYSING AND INTERPRETING DATA

Answer BOTH questions.

5. (a) The length of time, x , between the planting and germination of a type of pepper seed is believed to have a normal distribution, with mean, μ , and standard deviation, σ . A random sample of 49 such seeds was planted under identical conditions and the length of time between planting and germination was measured. The following totals were recorded.

$$\sum x = 510, \sum x^2 = 5635$$

- (i) Calculate unbiased estimates for the population mean, μ .

.....
.....

[2 marks]

- (ii) Calculate unbiased estimates for the standard deviation, σ , of the length of time between planting and germination.

.....
.....
.....
.....

[3 marks]

- (iii) Construct a 94% confidence interval for the mean length of time it takes for this type of pepper seed to germinate.

.....
.....
.....
.....

[3 marks]

DO NOT WRITE IN THIS AREA

GO ON TO THE NEXT PAGE

- (iv) Sixty random samples of 49 pepper seeds are taken and a 95% confidence interval for μ is found for each sample. Determine the appropriate number of intervals that will contain the population mean, μ .

.....

[2 marks]

- (b) The mean of 150 observations of X , where $X \sim \text{Bin}(135, 0.36)$ is \bar{X} . Use the central limit theorem to state the distribution that is modelled by \bar{X} , giving the values of its parameters.

.....

[5 marks]

- (c) A manufacturer claims that the mean mass of a certain type of soap bar produced is 110 g. A random sample of 18 such soap bars gave a mean mass of 105 g and a standard deviation of 6.7 g. Test, at the 5% level of significance, the hypothesis that the manufacturer is overstating the mean.

- (i) State, in symbols, appropriate null and alternative hypotheses for the test.

.....

[3 marks]

- (ii) Determine the critical region(s) for the test.

.....

[3 marks]

GO ON TO THE NEXT PAGE

DO NOT WRITE IN THIS AREA

(iii) Calculate the value of the test statistic.

.....
.....
.....
.....

[2 marks]

(iv) State your conclusions, giving reasons.

.....
.....

[2 marks]

Total 25 marks

DO NOT WRITE IN THIS AREA

GO ON TO THE NEXT PAGE

6. (a) The following contingency table summarizes the responses of a random sample of 500 prospective voters to a survey on voter preference for three candidates, *P*, *Q* and *Z*, contesting a general election.

Age of Voter (years)	Candidate			Total
	<i>P</i>	<i>Q</i>	<i>Z</i>	
18–30	45	16	13	74
31–45	98	91	24	213
46–60	33	70	29	132
Over 60	26	12	43	81
Total	202	189	109	500

A χ^2 test is carried out to determine whether there is an association between the age of the voter and the candidate preference of the voter.

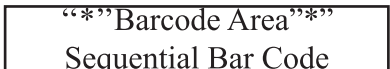
Formulate null and alternate hypotheses for this test.

.....

.....

.....

[2 marks]



(b) The number of voters in the survey who are over 60 years old and have a preference for Candidate P is shown in the cell corresponding to row 4, column 1.

(i) Calculate, for this cell, the expected number of voters.

.....
.....
.....

[2 marks]

(ii) Calculate the χ^2 value for this cell.

.....
.....
.....

[2 marks]

(iii) Determine the critical region of the test at the 5% level of significance.

.....
.....
.....
.....
.....

[3 marks]

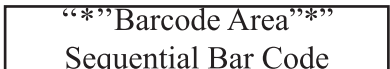
(iv) Write a valid conclusion for the test, given that the value of the test statistic is 92.195.

.....
.....
.....

[1 mark]

DO NOT WRITE IN THIS AREA

GO ON TO THE NEXT PAGE



- (c) A used car dealer is investigating the relationship between the price of a particular model of car and its age. The following information was collected from a sample of eight cars of this model.

Age of Car in Years (x)	8	3	6	9	2	5	6	3
Value of Car (\$1000) ($y$)	18	84	50	21	95	42	36	78

Summary: $\sum x = 42$; $\sum y = 424$; $\sum xy = 1\,735$; $\sum x^2 = 264$; $\sum y^2 = 28\,490$

The least squares regression line of y on x is given by the equation

$$y = 112.26 - 11.29x.$$

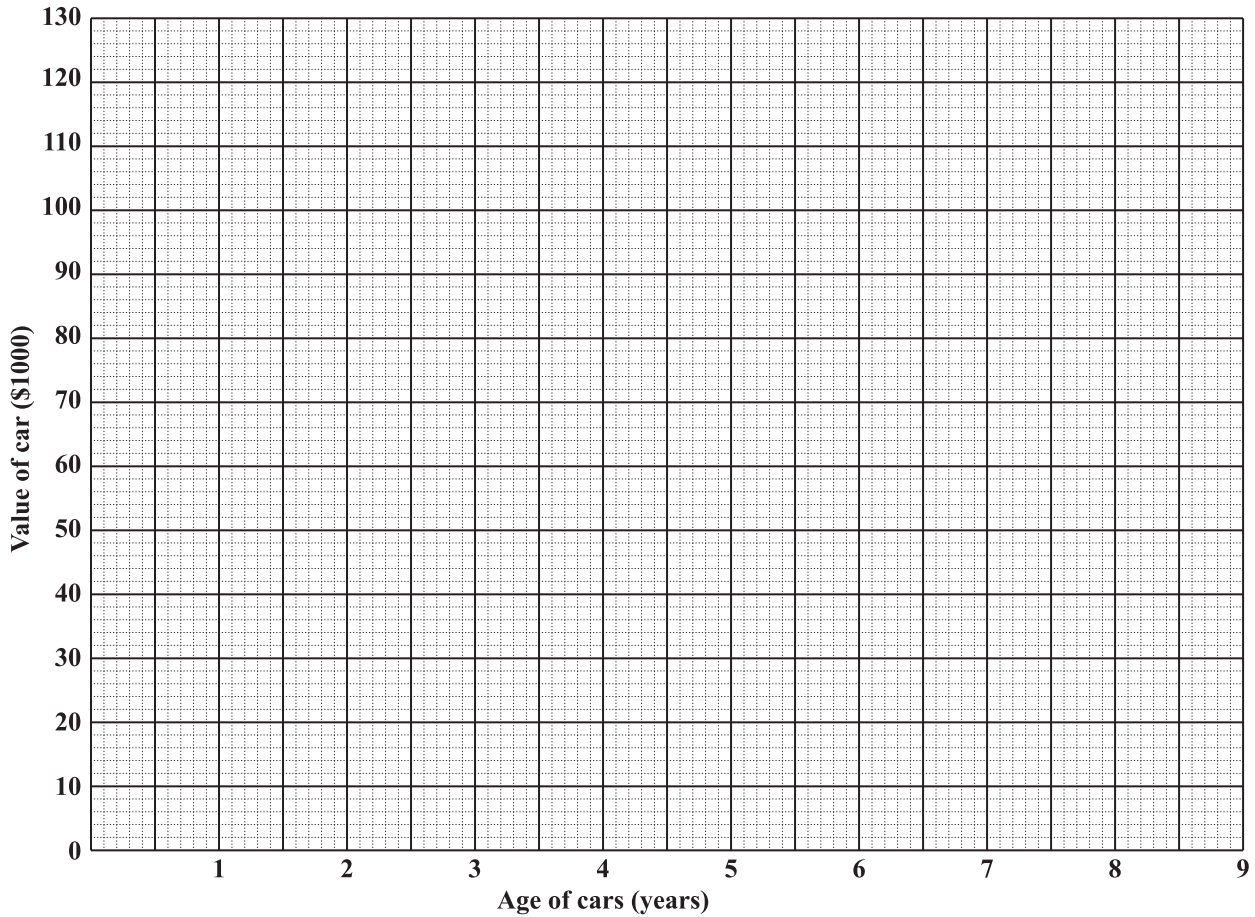
The Pearson's product-moment correlation coefficient is $r = -0.96$.

- (i) State ONE reason why the value of the car (y) is the dependent variable.

.....

[1 mark]

(ii) Using the grid below, plot the data on a scatter diagram.



[3 marks]

(iii) Calculate the mean value of the age of the car, (\bar{x}) , and the mean value of the value of the car, (\bar{y}) . Clearly show this point on the scatter diagram.

.....

.....

.....

.....

[3 marks]

GO ON TO THE NEXT PAGE

DO NOT WRITE IN THIS AREA

- (iv) On the same grid as the scatter diagram, draw the regression line

$$y = 112.26 - 11.29x.$$

[2 marks]

- (v) With respect to the data, explain the value -11.29 in the regression equation.

.....
.....
.....

[2 marks]

- (vi) Use the regression line to estimate the value of a car which is 7 years old.

.....
.....
.....
.....

[2 marks]

- (vii) Interpret the value of the product-moment correlation coefficient, r , as it relates to the data.

.....
.....
.....

[2 marks]

Total 25 marks

END OF TEST

IF YOU FINISH BEFORE TIME IS CALLED, CHECK YOUR WORK ON THIS TEST.

DO NOT WRITE IN THIS AREA