

# WATERFORD INSTITUTE OF TECHNOLOGY

## HIGHER CERTIFICATE IN COMPUTER APPLICATIONS

### EXAMINATION:

### STATISTICAL ANALYSIS YEAR 1

### (SEMESTER TWO EXAM)

**Summer 2007**

**DURATION: 2 HOURS**

**EXAMINER: MS. ANN VEREKER  
MR. JONATHAN BRAZIL**

### INSTRUCTIONS TO CANDIDATES

1. ATTEMPT ALL QUESTIONS
2. TOTAL MARK = 100
3. LINEAR REGRESSION FORMULAE AND STATISTICS FORMULAE ARE APPENDED FOR USE WITH QUESTIONS AS APPROPRIATE

### EQUIPMENT/MATERIALS SUPPLIED

1. GRAPH PAPER.

### Question 1

The following is a table of raw data that represents the age of a certain car in years and its average number miles per gallon. Suspicion suggests that the older the car is the less economic it is. Answer the questions that follow.

Age of car in years	8	3	6	9	2	5	6	3
Miles per gallon	12	14	10	9	15	12	9	15

- (i) Calculate the regression line equation. **(15 marks)**
- (ii) Draw the scatter diagram of this data. **(2 marks)**
- (iii) Draw the regression line using labelled axes. **(3 marks)**
- (iv) Calculate the coefficient of determination ( $r^2$ ). **(10 marks)**
- (v) If X denotes the age of a car selected at random and Y denotes the economy of that car in miles per gallon then explain the significance of the coefficient of determination. **(5 marks)**
- (vi) How many miles per gallon would you expect to get from:
- a) A 7-year-old car?
  - b) An 11-year-old car?
  - c) A 20-year-old car?

Explain your answers and their reliability.

**(5 marks)**

**(Total 40 marks)**

## Question 2

(i) What is the difference between discrete and continuous data? (2 marks)

(ii) Describe the Simple Random Sampling Method and provide a description of the problems associated with using this method to pick sample frames. (3 marks)

(iii) The following table represents the time endured by a class of students in a fitness test. Inspect the data and answer the questions that follow:

Time Endured (minutes)	No of students
10 – 15	2
15 – 20	3
20 – 25	6
25 – 30	8
30 – 35	4
35 – 40	2
40 – 45	1

Note: The interval (10 – 15) ranges from 10.0 minutes to 14.9 minutes

a) Present the data shown above in a cumulative frequency distribution table. (2 marks)

b) Draw an ogive of the data. (3 marks)

c) Draw a histogram of the data set. (3 marks)

d) Calculate the median of the data set. (6 marks)

e) Calculate the mode of the data set. (5 marks)

f) Calculate the inter-quartile range of the data set. (6 marks)

**(Total 30 marks)**

### Question 3

- (i) The following are the sales figures for a company over a twelve-month period in 2006.

JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
92	84	73	61	63	72	73	80	78	90	85	99

Calculate the moving averages forecasts with  $n=4$  from May 2006 to January 2007.

**(10 marks)**

- (ii) The table below represents the sales of a particular item over a 3-year period.

Year	Jan-Mar	Apr-Jun	Jul-Sep	Oct-Dec
2004	250	290	310	400
2005	220	280	340	410
2006	240	320	280	390

- a) Given that the regression line, plotting sales vs. time, is given by  $y=258.6+9.5x$ , determine the seasonally adjusted forecasts for each quarter in the above table.

**(18 marks)**

- b) Determine the seasonally adjusted forecasts for each quarter in 2007.

**(2 marks)**

**(Total 30 marks)**

## Statistical Formulae

1. Quartiles for grouped data

$$Q = L + \frac{r}{f}(U - L)$$

2. Mode for grouped data

$$Mode = L + \frac{d_1}{d_1 + d_2}(U - L)$$

## Linear Regression

1. Linear parameters a and b

$$S_{xy} = \frac{\sum xy}{n} - (\bar{x})(\bar{y})$$

$$S_{xx} = \frac{\sum x^2}{n} - (\bar{x})^2$$

$$S_{yy} = \frac{\sum y^2}{n} - (\bar{y})^2$$

$$b = \frac{S_{xy}}{S_{xx}} \quad a = \bar{y} - b\bar{x}$$

2. Correlation coefficient

$$r^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$