

WATERFORD INSTITUTE OF TECHNOLOGY

HIGHER CERTIFICATE IN COMPUTER APPLICATIONS

EXAMINATION:

STATISTICAL ANALYSIS YEAR 1

(SEMESTER TWO EXAM)

Summer 2006

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

A project manager in a software development company wishes to investigate if the number of programming errors a computer programmer makes in every 500 lines of code written reduces as the programmer gains more work experience. The project manager collects the following data from a sample of eight programmers.

No. of Programming Errors (Per 500 Lines of Code)	15	9	5	19	12	1	26	1
No. of Yrs Working (As a Programmer)	10	18	24	5	9	30	2	29

- (a) Find the least squares regression line. **(16 marks)**
- (b) Give a brief description of the values for a and b in the regression line equation calculated in part (a). **(8 marks)**
- (c) Plot the above data and the regression line on the same set of axes. **(8 marks)**
- (d) Calculate the correlation coefficient and comment on its value. **(8 marks)**

(Total 40marks)

Question 2

(a) Give a brief description of each of the following data types:

- (i) Nominal
- (ii) Ordinal
- (iii) Cardinal

(6 marks)

(b) Describe each of the following sampling methods:

- (i) Simple Random Sampling Method
- (ii) Stratified Sampling Method.

(4 marks)

(c) The following table gives the results of students in mathematics exam.

35	47	93	21	54
36	15	74	18	9
56	81	32	45	71
63	41	61	91	83
33	55	62	83	24

Using the above data:

- (i) Present the data in a frequency distribution table.

(2 marks)

- (ii) Present the data in a cumulative frequency distribution table.

(2 marks)

- (iii) Draw a Histogram and Ogive to represent the above data.

(4 marks)

- (iv) Calculate the mean and the median.

(6 marks)

- (v) Calculate the standard deviation and the range.

(6 marks)

(Total 30 marks)

Question 3

- (a) The following are the sales figures for a company over a twelve month period in 2005.

JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC
100	95	82	70	61	52	48	39	45	69	78	94

Calculate the moving averages forecasts with $n = 3$ from April 2005 to January 2006.

(10 marks)

- (b) Below is tabulated the sales of a particular item for each quarter over a 3 year period.

Year	Spring	Summer	Autumn	Winter
2003	37	40	45	53
2004	62	70	77	85
2005	93	100	111	127

- (i) Given the regression line plotting sales with time is given by $Y = 40 + 5X$. Determine the seasonally adjusted forecasts for each season in the above table.
- (ii) Determine the seasonally adjusted forecast for the first season in 2006.

(18 marks)

(2 marks)

(Total 30 marks)

Statistical Formulae.

1. Quartiles

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

2. Arithmetic mean for grouped data

$$\bar{x} = \frac{\sum fx}{\sum f}$$

3. Standard deviation for grouped data

$$\sigma = \sqrt{\frac{\sum f(x - \bar{x})^2}{\sum f}}$$

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.

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