S

93201Q

NEW ZEALAND QUALIFICATIONS AUTHORITY
MANA TOHU MĀTAURANGA O AOTEAROA

Scholarship 2005
Statistics and Modelling

9.30 am Monday 5 December 2005
Time allowed: Three hours
Total marks: 156

QUESTION BOOKLET

You should complete ALL the questions.

You should write ALL your answers in the Answer Booklet 93201A.

Show ALL working. Start each question on a NEW page. Number each question carefully.

A Mathematical Formulae and Tables Booklet S-STATF is provided in the centre of this examination paper.

Check that this question booklet has pages 2–9 in the correct order, as well as a centre pull-out of 4 pages of mathematical formulae and tables, and that none of these pages is blank.

The approximate time you should spend on each question is as follows:

Question	Marks	Time (minutes)
One	22	35
Two	28	25
Three	26	20
Four	46	40
Five	14	20
Six	20	40

YOU MAY KEEP THIS BOOKLET AT THE END OF THE EXAMINATION.

All questions are about a business named *Megabitz,* which operates as three different modes. It has a computer store, internet café and a refreshment bar.

QUESTION ONE (22 marks)

The refreshment bar sells two types of pie, steak and mince. The number of pies ordered each day needs to satisfy the following daily constraints.

1. At least 42 pies will be sold in total.
2. A minimum of 24 mince pies will be sold.
3. No more than 3 mince pies will be sold for every steak pie sold.
4. At most, 12 more steak pies will be sold than mince pies.

The refreshment bar pays $1.30 for each steak pie and $1.00 for each mince pie.

(a) Find the number of steak pies and mince pies that should be ordered each day in order to minimise the daily cost of buying the pies. (6 marks)

(b) Assume the cost of steak pies remains at $1.30 each. The cost of mince pies increases so that the optimal solution is no longer that obtained in (a). State the possible changes in the cost of mince pies that give rise to these different optimal solution(s), and give the new optimal solution(s) for each of these changes. (8 marks)

(c) Suppose the demand for the pies is modelled by the function: $S^2 = 66M + 10$ where

S represents the daily demand for steak pies, and
M represents the daily demand for mince pies.

Assume that the refreshment bar pays $1.30 for each steak pie and $1.00 for each mince pie.

Find the optimal solution that satisfies the daily demand constraints and the demand function, and minimises the daily cost of buying the pies. (8 marks)

QUESTION TWO (28 marks)

The computer store sells two types, A and B, of a circuit component that cannot be distinguished by eye. However, the components have varying weights, and overall, type B components are heavier than type A components. The weights (in grams) of each type can be modelled by a normal distribution with parameters as follows:

Parameter	Type A	Type B
Mean	22.6	27.2
Standard Deviation	1.2	1.0

As a result the following system is suggested for determining if a component is of type A or type B.

- Weigh the component.
- If the weight is less than 24.9 g, the component is classified as type A, otherwise it is classified as type B.

Type A and type B components occur randomly.

(a) 1. What percentage of type A components will be wrongly classified as type B? (6 marks)

2. Assuming there will be equal proportions of type A and type B components, what percentage of incorrectly identified components will be of type B? (8 marks)

3. Suppose that the components come in bags of 20. Also assume that all components in a bag are of the same type but the type cannot be determined by eye.

Suggest an improvement to the classification system that will decrease the percentage of incorrectly classified components.

Justify your answer. Include at least one calculation in your justification. (8 marks)

(b) Customers arrive at the store at a constant average rate of 12 per hour. Assuming arrivals are random and independent, find the number of customer arrivals in a half-hour interval that, in the long run, would be exceeded only in 5% of half-hour intervals. Explain your answer fully. (6 marks)

QUESTION THREE (26 marks)

(a) Based on a random sample of 20 *Megabitz* customers, a 95% confidence interval for the mean age (in years) of *Megabitz* customers is (24.9015, 31.7385).

Using this information, calculate a 99% confidence interval for the mean age (in years) of *Megabitz* customers. (6 marks)

(b) Calculate how many more *Megabitz* customers should be sampled in order to estimate the mean age to within two years with 99% confidence. State any assumption(s) made. (8 marks)

(c) It was desired to compare the mean expenditures of male and female *Megabitz* customers. Two independent random samples of 30 male and 40 female customers gave the following summary statistics:

Summary Statistics	Males	Females
Sample mean	\$10.02	\$11.13
Sample standard deviation	\$1.81	\$1.59
Sample size	30	40

1. Construct a 95% confidence interval for the difference between the population mean expenditures for male and female *Megabitz* customers. (6 marks)

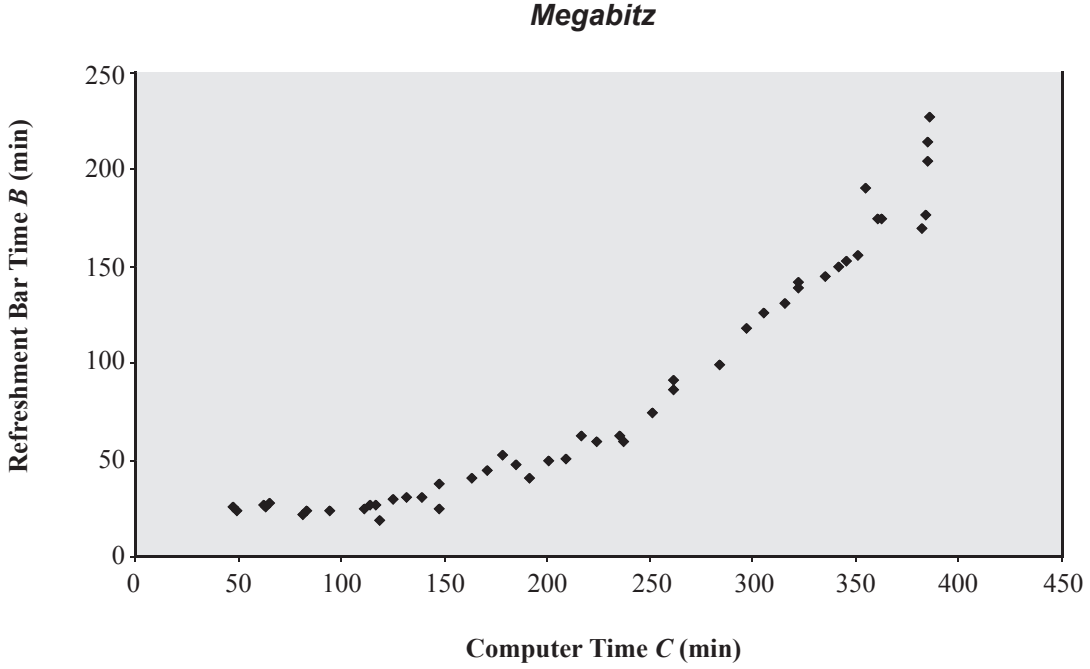
2. Is there a difference between the population mean expenditures for male and female *Megabitz* customers? Justify your answer. (6 marks)

QUESTION FOUR (46 marks)

The owner of *Megabitz* wishes to create surroundings to suit the customers who use the internet café. She wishes to determine whether there is any relationship between the time that people spend on the computers and the time that these people spend at the refreshment bar.

A random sample of 49 customers was used to determine a model for predicting the time spent at the refreshment bar from the time spent on the computers.

Let B represent the time spent at the refreshment bar (in minutes) and C represent the time spent on the computers (in minutes). A scatter plot of B against C is shown opposite at the top of page 5:



Two models are investigated.

Model 1
A linear regression model gives a fitted line with equation: $B = -36.332 + 0.5477C$.
For this line $R^2 = 0.8951$.

Model 2
A log (base e) transformation of variable B is carried out. A linear regression model gives a fitted line with equation: $\ln B = 2.5810 + 0.0071C$.
For this line $R^2 = 0.9627$.

(a) By reference to the scatter plot, fully describe the relationship between B and C. (6 marks)

(b) For Model 2, by first making B the subject, sketch B versus C to show the general shape of this function for $0 \leq C \leq 400$. (6 marks)

(c) Which of these models is likely to be better for predicting time spent at the refreshment bar for low users (less than 30 minutes) of computers? Justify your choice. (6 marks)

(d) Using both Model 1 and Model 2, obtain two predictions, in each case, of the time spent at the refreshment bar by a person who spends on the computers: (6 marks)

- three hours
- six hours 45 minutes.

(e) Comment on the reliability of your predictions in part (d). (8 marks)

(f) Calculate the range of C values for which Model 1 gives higher predictions than Model 2. (6 marks)

(g) Discuss the appropriateness of using Model 2 to model the overall relationship between B and C in comparison to Model 1. (8 marks)

QUESTION FIVE (14 marks)

In an effort to investigate the relationship between the age of customers using the refreshment bar and their expenditures, the owner takes a sample of 60 customers from three age groupings A, B and C in the ratio 3:1:2 respectively. These age groupings are as follows:

- A: 15–25 years
- B: 26–34 years
- C: 35–45 years.

The owner's samples from each of these age groupings A, B and C yielded total expenditures of \$444, \$206 and \$306 respectively.

The owner decides to model average expenditure with a quadratic function $y = ax^2 + bx + c$, where x represents the mid-point of the age grouping (in years) and y represents average expenditure (in dollars) for the corresponding age grouping.

(a) Form a system of equations and solve it to find this quadratic function model used to describe average expenditure in terms of age. (6 marks)

(b) Based on this model, the owner decides that it is not worth creating a special lounge area in her refreshment bar for people older than 50. Comment on this decision and justify your comment. Include at least one calculation in your justification. (8 marks)

Note that QUESTION SIX is on Page 8.

QUESTION SIX (20 marks)

You are contracted as a statistical analyst to investigate sales patterns for the internet café over the previous three years, and to make a sales forecast for February 2006.

You have been provided with the following data:

The value of sales ($000) from the café for each month from November 2002 ($t = 1$) to October 2005 ($t = 36$) inclusive.

The 12-point centred moving average (CMA) sales values ($000) for the months May 2003 ($x = 1$) to April 2005 ($x = 24$) inclusive.

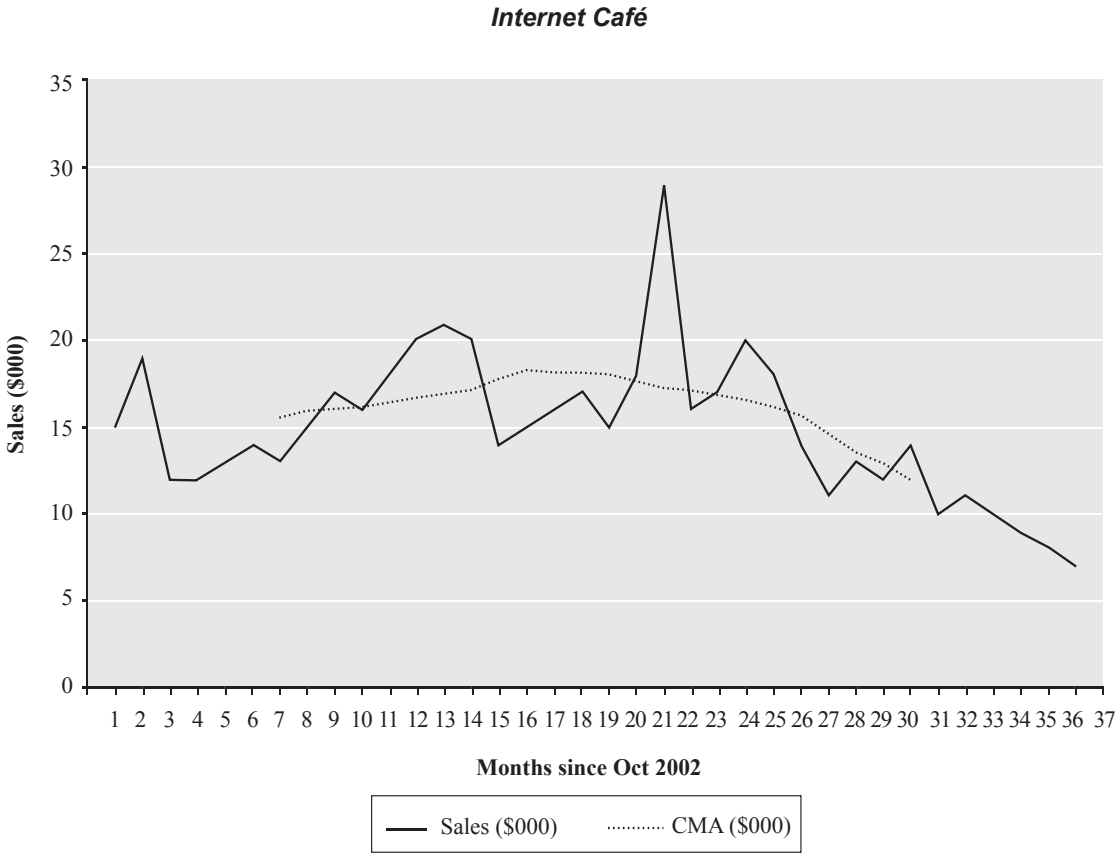
In addition you are provided with the following statistical output or information:

- A graph of the value of monthly sales on which the CMA values have also been plotted.
- A table showing some summary statistics for the sales over each six-monthly period.
- A linear regression line fitted to the plotted CMA points has equation: $y = -0.1112x + 17.709$ and $R^2 = 0.2249$.
- A quadratic regression curve fitted to the plotted CMA points has the equation: $y = -0.0324x^2 + 0.6987x + 14.199$ and $R^2 = 0.9527$.

Write a report, no more than a page long (excluding calculations), to the owner of *Megabitz* that summarises the output.

Include in your report two calculations, one using the line and the other using the curve, to forecast the sales for February 2006.

Comment on the usefulness and limitations of your forecasts.



Summary of Monthly Sales ($000) in each Six-monthly Period from Nov 2002

Summary Statistics	Nov '02 – Apr '03	May '03 – Oct '03	Nov '03 – Apr '04	May '04 – Oct '04	Nov '04 – Apr '05	May '05 – Oct '05
Mean	14.17	16.50	17.17	19.17	13.67	9.17
Median	13.50	16.50	16.50	17.50	13.50	9.50
Standard Deviation	2.64	2.43	2.79	5.12	2.42	1.47

***t*-value**	**Month**	**Year**	**Sales ($000)**	***x*-value**	**CMA ($000)**
1	Nov	2002	15		
2	Dec	2002	19		
3	Jan	2003	12		
4	Feb	2003	12		
5	Mar	2003	13		
6	Apr	2003	14		
7	May	2003	13	1	15.58
8	June	2003	15	2	15.88
9	July	2003	17	3	16.00
10	Aug	2003	16	4	16.21
11	Sept	2003	18	5	16.46
12	Oct	2003	20	6	16.71
13	Nov	2003	21	7	16.92
14	Dec	2003	20	8	17.13
15	Jan	2004	14	9	17.75
16	Feb	2004	15	10	18.25
17	Mar	2004	16	11	18.21
18	Apr	2004	17	12	18.17
19	May	2004	15	13	18.04
20	June	2004	18	14	17.67
21	July	2004	29	15	17.29
22	Aug	2004	16	16	17.08
23	Sept	2004	17	17	16.83
24	Oct	2004	20	18	16.54
25	Nov	2004	18	19	16.21
26	Dec	2004	14	20	15.71
27	Jan	2005	11	21	14.63
28	Feb	2005	13	22	13.54
29	Mar	2005	12	23	12.88
30	Apr	2005	14	24	11.96
31	May	2005	10		
32	June	2005	11		
33	July	2005	10		
34	Aug	2005	9		
35	Sept	2005	8		
36	Oct	2005	7		