S

93201Q

932012

NEW ZEALAND QUALIFICATIONS AUTHORITY
MANA TOHU MĀTAURANGA O AOTEAROA

Scholarship 2007
Statistics and Modelling

2.00 pm Tuesday 20 November 2007
Time allowed: Three hours
Total marks: 48

QUESTION BOOKLET

A 4-page booklet (S-STATF) containing mathematical formulae and tables has been centre-stapled in the middle of this booklet. Before commencing, carefully detach the Formulae and Tables Booklet and check that none of its pages is blank.

Answer ALL questions.

Write ALL your answers in the Answer Booklet 93201A.

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

Check that this Question Booklet 93201Q has pages 2–10 in the correct order and that none of these pages is blank.

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

This page has been deliberately left blank.

You have three hours to complete this examination.

All of the questions are based on the packing of kiwifruit into trays at a packing shed called Shedz. A number of growers send their kiwifruit to Shedz for packing.

QUESTION ONE (8 marks)

(a) (i) Shedz processes two varieties of kiwifruit – green and gold. Constraints on daily packing are as follows:

- At most, 2 000 trays can be packed in total.
- At most, 1 200 trays of green can be packed.
- A minimum of 200 trays of gold need to be packed.
- The ratio of the number of packed trays of green to gold must be no more than 5:2.

The ratio of the profit per tray of the green variety to the gold is 12:13.

Find the number of trays of each variety that should be packed each day to maximise the profit. **You may use the graph paper provided on page 2 of your Answer Booklet.**

(ii) The output of Shedz should reflect market demand and maintain a strong profit.

Given that there is a market demand for both green and gold varieties, suggest what other solutions to part (a) (i) are possible that do not involve changing any constraints. Justify your suggestions fully.

(b) Three types of tray are available; "all cardboard", "cardboard / wood" and "all wood", and Shedz wishes to determine how many of each type to purchase.

The cost per tray is \$1 for "all cardboard", \$2 for "cardboard / wood" and \$4 for "all wood". Shedz budgets \$5 000 per day for purchasing trays. Shedz requires 2 000 trays per day, with between 10% and 30% of them being of the "all cardboard" type and as many as possible (subject to the above criteria) being of the "cardboard / wood" type.

Using these criteria, determine Shedz's daily requirement for each type of tray.

QUESTION TWO (8 marks)

(a) Shedz buys its trays in pallets by the container load. As each container arrives, n trays are examined to assess whether or not the load of trays is of acceptable quality. The load is accepted if there is no more than one defective tray in the sample.

(i) For two successive loads, 1% and 2% of the trays were defective, respectively.

If $n = 50$ for each load, find the probability that at least one of these two loads was accepted.

(ii) If 1% of the trays are defective in a load, find the largest value of n which ensures there is approximately a 94% chance of the load being accepted.

(b) Each pallet consists of 180 trays stacked in 30 levels, with six trays arranged as a 3 by 2 rectangle at each level.

(i) Describe how cluster sampling could be used to select 24 trays from a container load of 20 pallets.

(ii) Give one advantage and one disadvantage of using cluster sampling in this situation, compared with simple random sampling.

QUESTION THREE (8 marks)

(a) The weight of a single kiwifruit and the weight of a tray with packaging vary with the following parameters:

	Mean	**Standard Deviation**
Single Fruit	105.6 grams	4.2 grams
Tray and Packaging	499.7 grams	19.7 grams

Find the mean and standard deviation of the weight of a packed tray of 36 kiwifruit. What assumption needs to be made?

(b) In order to monitor the production from each grower, a random sample of six packed trays is taken periodically. The weight of each packed tray is found, and the sample mean calculated.

(i) Shedz has set limits, equally spaced either side of the population mean, within which the mean weight of six packed trays would be expected to lie for 95% of all samples.

Assuming a normal distribution for the weights of a packed tray and using your answer to (a), find these limits.

(ii) What would you conclude if the total weight of a sample of six packed trays from a particular grower was 26 kilograms? Justify your answer.

(c) Using your previous answers, find the sample size taken when the probability of the sample mean being within 0.005 kg of the population mean is 0.89, given that this sample mean is within the 95% limits equally spaced either side of the population mean for this sample size.

QUESTION FOUR (8 marks)

Some kiwifruit have marks that are caused by rubbing between adjacent kiwifruit while they are on the vine. Shedz wants to investigate whether the size of the mark is related to the weight of the kiwifruit. A random sample of 26 kiwifruit with such a defect was selected. The weight, W (in grams), of each kiwifruit was found and the size, S (in millimeters), of each mark was found by measuring its greatest width.

Table 1 shows the data, Figure 1 shows a scatter plot of this data, and Figure 2 shows a residual plot of this data.

Table 1

W (g)	S (mm)	W (g)	S (mm)
106	4.5	110	0.9
110	5.4	100	2.5
110	6.1	100	4.2
102	4.4	108	4.9
101	3.5	109	5.3
111	6.3	113	6.3
106	4.7	103	3.5
98	2.5	112	1.4
98	1.5	111	5.9
99	3.4	106	4.2
109	5.1	111	1.2
103	4.1	104	3.5
99	1.6	109	5.7

Figure 1

Kiwifruit Quality

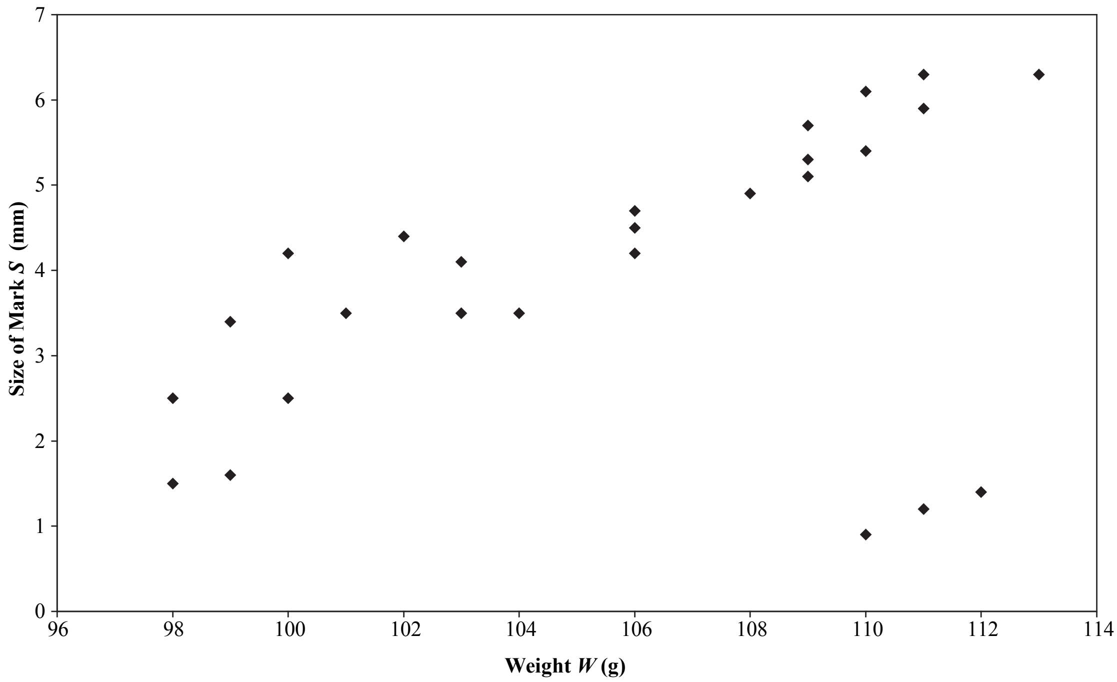
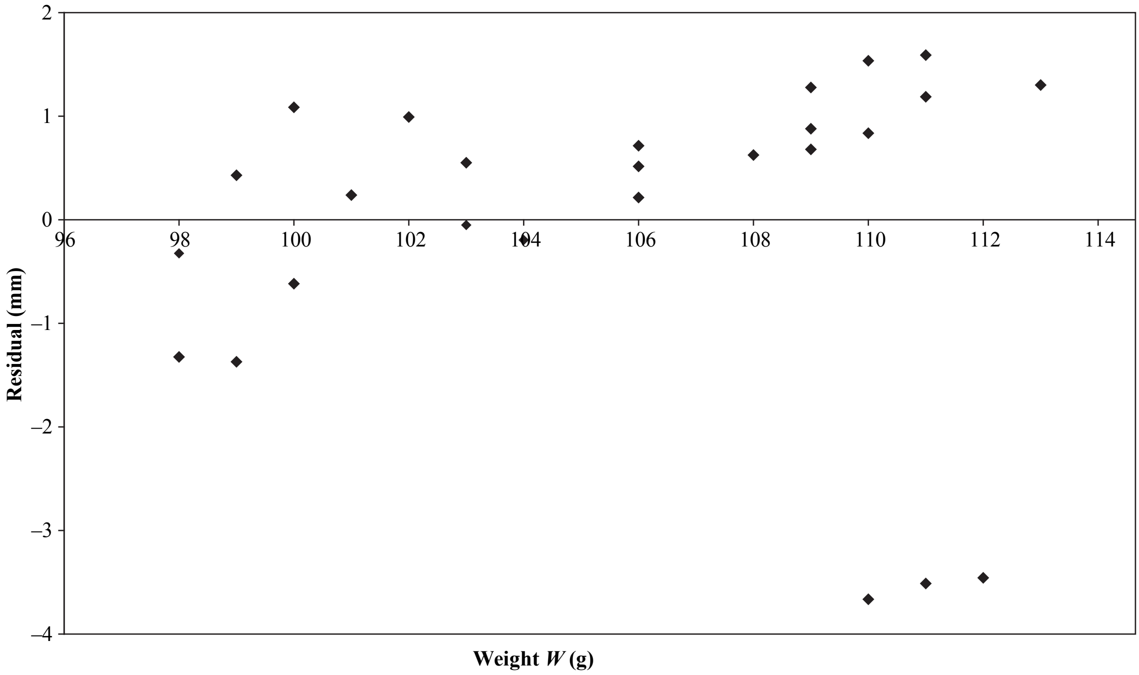


Figure 2
Kiwifruit Quality



Two regression lines were fitted to the data, one of them to the complete set of 26 observations, and the other to 23 observations. The equations of the regression lines (not necessarily respectively) are:

Line A: $S = 0.2724W - 24.293$ with $R^2 = 0.8514$, and

Line B: $S = 0.1450W - 11.384$ with $R^2 = 0.1798$

(a) Write a short paragraph to describe the relationship between the weight of a kiwifruit and the size of a mark. Include at least four features of the relationship.

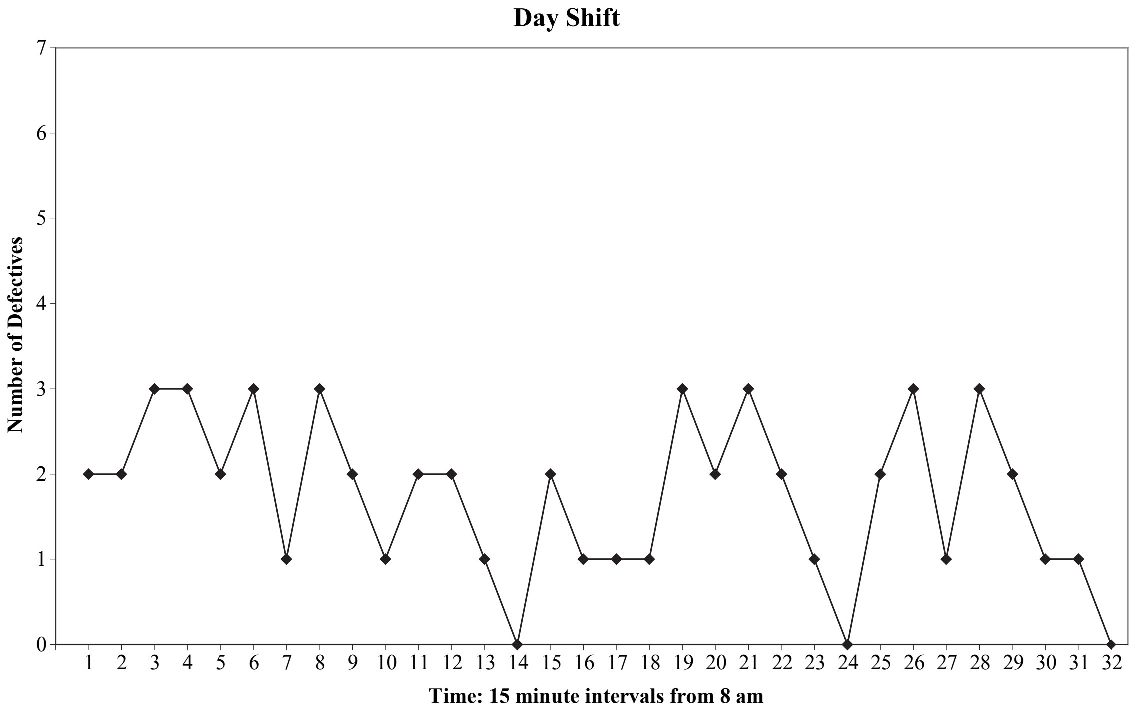
(b) By making an appropriate choice of regression line from either A or B, predict the size of a mark for kiwifruit with weights of 89 g and 115 g. Justify your choice of line. Also comment on the validity of each prediction.

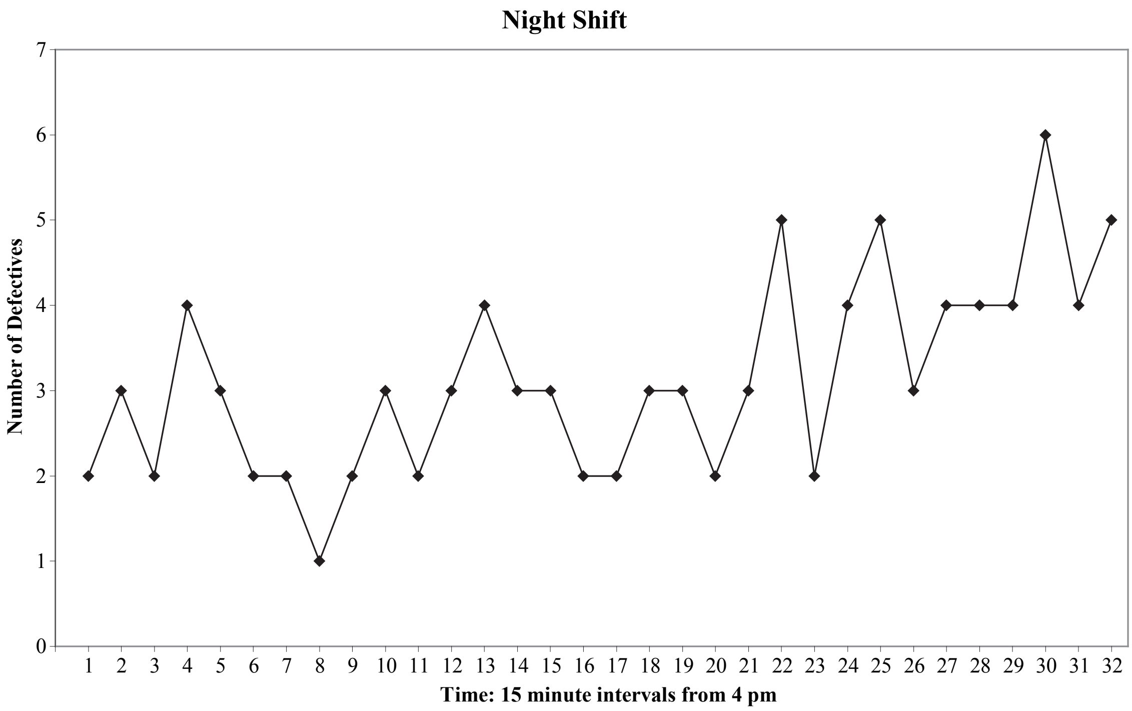
(c) Suggest an improvement to this investigation.

QUESTION FIVE (8 marks)

It is suspected that the identification of defective kiwifruit is affected by lighting, especially at night. A sampling strategy was designed to compare the number of defective kiwifruit found during one day shift (8 am to 4 pm) and one night shift (4 pm to midnight), where all the kiwifruit was from one grower. During packing, two trays (36 kiwifruit per tray) were taken every 15 minutes, and the number of defective kiwifruit in the total of 72 kiwifruit examined was recorded.

The results are displayed on the two plots below, along with summary statistics for the number of defective kiwifruit found in two trays at the top of the following page.





Summary Statistics for Number of Defective Kiwifruit

Statistic	**Shift**	
	Day	**Night**
Mean	1.75	3.13
Standard Deviation	0.95	1.16
Median	2	3
Range	3	5

(a) Write a one-page essay that compares the number of defective kiwifruit found during the day shift with that found during the night shift. Include a discussion of whether the percentage of defectives found in each shift is within a quality requirement of no more than 2.5% of the kiwifruit being defective.

(b) Describe how you would carry out a statistical test to see if the mean number of defective fruit found on night shifts is generally different from the mean number of defective kiwifruit found on day shifts. Include some details on how you would sample, and state any assumption(s) that you would need to make.

QUESTION SIX (8 marks)

The quality of the kiwifruit packed on any day is assessed by taking a random sample of n kiwifruit and counting the number of defective kiwifruit, d, in the sample. Initially 20 kiwifruit are selected (ie $n = 20$) and the value of d is recorded. The following decision criteria are then applied:

Step 1: If $d > 0.05n + 1$, all the kiwifruit packed on that day are rejected.
If $d \leq 0.05n + 1$, go to Step 2.

Step 2: If $d < 0.37\,e^{0.04n} - 1$, all the kiwifruit packed on that day are accepted.
If $d \geq 0.37\,e^{0.04n} - 1$, go to Step 3.

Step 3: A further sample of 5 kiwifruit is taken.
n is reset to the cumulative sample size (ie, n takes values 25, 30, 35,...), and d is reset to the cumulative number of defective kiwifruit.
Go to Step 1.

These criteria continue to be applied until all the kiwifruit packed on that day are either accepted or rejected.

(a) (i) Sketch the graphs of $d = 0.05n + 1$ and $d = 0.37\,e^{0.04n} - 1$ for $0 \leq n \leq 80$ on the same set of axes showing all intercepts. **Use the graph paper provided on page 26 of your Answer Booklet.**

(ii) Find the maximum value of n at which a decision to accept or reject must occur and, for this value of n, find the minimum value of d that would result in a rejection decision.

(b) Find the probability that kiwifruit packed on a day is accepted at $n = 25$, given that it contains 4% defective kiwifruit.

(c) If the initial sample of 20 has 2 defective kiwifruit, would you conclude that the proportion of defective kiwifruit packed on that day is different from 2.5% (the maximum tolerance for defective kiwifruit)? Justify your answer by performing at least one calculation.

93201Q