

Assessment Report

Scholarship, 2007

Statistics and Modelling

Statistics and Modelling, Scholarship, 2007

Commentary:

On the whole questions 1(b), 2(a) (i), Q2 (b) (ii), 3(a), 4(a), 5(a) were well done. The questions found hardest by the candidates were Q3(c) and Q6 (b).

Some candidates were ill-prepared to answer questions that had content from level 1 and 2 achievement standards. Full evidence of steps taken to reach answers like probability statements were not given, especially when candidates used graphics calculators. Candidates must write legibly. Some handwriting was almost too difficult to read.

The best-performing candidates most commonly demonstrated the following skills and / or knowledge:

- effective communication of statistical ideas with clarity and efficiency of observations. Their answers were set out in a clear and logical way with very good layout, presentation, and writing. They were able to write well-structured paragraphs and essays
- a high level of advanced statistical thinking
- ability to plan statistical answers and draw appropriate conclusions in unfamiliar situations
- ability to apply probability theory and mathematical knowledge to solve complex problems in the construction of confidence intervals and the calculation of probabilities
- identification of steps required for a statistical test along with the assumptions required to carry out the test in the context of the problem
- ability to write equations and constraints to model situations
- ability to read the questions and interpret them with well-written and nonrepetitive answers identifying the key features of the data and graphs provided
- being able to focus on the requirements of questions and correctly interpret them
- correctly justify the choice of model and discuss the validity of the resulting predictions.

Candidates who did not achieve Scholarship lacked some or all of the skills and knowledge above and, in addition, they:

- made vague, superficial, and irrelevant comments, often repeating the same feature in different words
- could not describe information contained in a graph
- had difficulty calculating the standard deviation
- calculated the confidence interval for six trays by taking six times the confidence interval for one tray
- could not relate their answer to the context of the question
- regurgitated the same point several times
- ignored in their discussions statistics given in the question
- were unable to correctly identify and discuss features of a residual plot
- incorrectly stated that a larger sample will improve the model
- outlined improvements to the model rather than improvements to the investigation
- were unable to select and apply knowledge and skills from separate standards to solve problems showing a limited appreciation of the complexity involved
- could not use correct terminology, eg stating that the size of the mark is proportional to the weight of the fruit.

Specific comments about the questions:

Question One

This question was well done by those who could define the variables and interpret the ratio type constraints when written in a sentence. Some candidates had difficulty with “the ratio must be no more than 5:2”, and in understanding the objective function so the optimal point was not found.

Some candidates:

- failed to show all the constraints.
- did not really understand the difference between gradient (sliding line) and optimal solution.

Question Two

Cluster sampling was not clearly described. The Binomial distribution was set up correctly; however, some candidates could not proceed from there by just looking in the tables.

Question Three

Common errors by candidates were:

- limits were given for the total rather than the mean
- standard error for the total was calculated incorrectly
- conditional probability was not recognised in (c)
- not getting the standard deviation correct
- not identifying “independence of the weights of the kiwifruit” as an assumption.

Very few got question 3(c) correct; in fact, most candidates had no idea where to start.

Question Four

Many candidates gave improvements to the model rather than the investigation.

Key words positive and linear were required to describe the correlation. The pattern in residual plot was missed by most. Candidates were worried by the negative answer to line A so incorrectly switched to Line B.

“Take more samples” was a common answer for part (c). Also, many said they wanted to exclude the outliers rather than investigate the reason for those outliers.

Question Five

Some candidates with graphical calculators were unable to change the scale in order to see the required domain and range.

The word “compare” was not appreciated in the answers ie both shifts were being compared. Three pages were written instead of the one page asked for. Statistics were simply listed for both day and night shifts with no actual comparison being made. Very few candidates made a comparison of trend of constant versus increasing.

Details were missing in the description of the use of the confidence interval for differences in the mean for the comparison of shifts.

Question Six

The values of n and d were often not given as whole numbers and that n progressed in fives.

Many candidates recognised that confidence intervals would be a good approach; however, they were unable to write a complete answer.

In (b), $n = 25$ and $\pi = 0.04$ were used and very few candidates went through the $n = 20$ process correctly.

A lot of candidates compared 2.5% to 10% and took no account of variation and didn't realise that a confidence interval was required.