ROYAL STATISTICAL SOCIETY EXAMINATIONS, 2008

REPORTS OF EXAMINERS

General Comments

Most comments made by examiners refer to specific features of questions set in this year's papers. But every year examiners draw attention to general aspects of examination technique that could be improved. As we have noted in earlier reports, it is disappointing to see candidates losing marks unnecessarily. Several comments made by examiners in 2008 echo those made in recent years. We therefore repeat here the advice to candidates given at the start of previous years' reports, revised so as to incorporate further general comments made by examiners following the 2008 papers.

The published syllabuses for the Higher Certificate and Graduate Diploma give details of mathematical topics with which candidates at those levels are expected to be familiar before embarking on study for the Society's examinations. You must make yourself aware of the necessary mathematics background for the examinations you intend to sit, and make every effort to master it.

Read any question you intend to answer slowly and carefully, and ensure you answer the question actually asked. Every year, some candidates reproduce bookwork that may have some relation to the topic but does not answer the question itself. Examiners award marks in accordance with detailed marking schemes, which assign marks for specific answers to each part of each question. There is therefore no point in writing down what you know about a different (if similar) topic, since the marking scheme will have no marks available for this.

On a related matter, be sure to carry out any specific instructions given in a question: e.g. round answers to three significant figures if that is what is asked; calculate the standard deviation, not just the variance, if that is what the question requires.

Take note of the marking scheme printed on the paper. It is a waste of your time writing a detailed two-page description of some topic, if this can only be awarded two marks.

When preparing for an examination, you will of course know that there will be certain details (definitions, formulae and the like) that you will be expected to have memorised. For any paper, candidates will be expected to know the definitions of all concepts relevant to the syllabus. As for formulae, it will be clear that (for example) a candidate who does not know the formula for a binomial probability function cannot fully understand the binomial distribution, so examiners may expect candidates to be able to quote that probability function when it is relevant to a syllabus. Similar examples can be given for other areas: formulae for sample variance and conditional probability (at Ordinary Certificate level) and sums of squares for appropriate analysis of variance models (at higher levels).

Make sure you understand the difference between the instructions *explain* and *define*. An *explanation* of some concept requires one or more sentences; the concept concerned should be described in words and (if appropriate) the purpose or use should be outlined. In a mathematical examination, a *definition* is a short and precise statement, which may require the use of mathematical notation. If a definition is required, a rough description is likely to be awarded no marks.

Ensure that you include sufficient reasoning in your answers for the examiners to be sure about the basis for any conclusions you draw. For example, writing "the test statistic is greater than the value in tables" without stating the value, the relevant sampling distribution or the degrees of freedom, will gain very few marks, if any.

In questions requiring calculations, it is understandable that errors will be made under examination conditions. When a candidate shows his or her working clearly, it is possible to give credit for use of a correct method even if there are errors in the numbers presented. However, when little or no working is shown, it is rarely possible to assess either the method being used or the source of the error. Candidates are therefore advised to show sufficient working to make it quite clear which method is being used.

When you complete a calculation, or finish answering a practical part of a question, try to check the plausibility of your result. For example, a variance cannot be negative, and a correlation coefficient cannot be outside the range –1 to +1. Similarly, a trend or regression line that does not pass through the main part of the data points is very unlikely to be correct.

If a rough sketch diagram is required, this can be done in your answer book; there is no need to draw it accurately on graph paper. This might for example apply to a sketch of a probability density function. Of course, such sketches must always be sufficiently clear that salient features stand out properly. However, when an accurate graph or chart is required, this should always be done on graph paper; and you should make sure you include a title and label the axes. This might for example apply to histograms.

It is important to follow the instructions on the front cover of the answer book. We realise that candidates will have little time to spend on reading the front cover during the examination itself, so we have produced a copy you can consult on the Society's website. You are strongly encouraged to look at this before the examination, and to ensure that you follow the instructions. We draw your attention to the following instructions in particular:

1. Begin each answer on a new page. (You do **not** need to begin each **section** of an answer on a new page.)

2. Write the number of each question at the top of each page.

3. Graph paper should be attached opposite the answer to which it relates.

4. Enter in the space below (**not** in the side panel) the numbers of the questions attempted. (The question numbers should be written **in the order in which you answered the questions**. Note that the side panel is for the examiners' use only.)

It is also helpful to examiners, as well as simpler for candidates, when the answer to a question is written on consecutive pages of the answer book. We do realise that, in practice, candidates may sometimes need to return to a question later. If you do this, it is helpful if you indicate this clearly on the page where the earlier attempt was made.

Ordinary Certificate

Ordinary Certificate Paper I

General

Most candidates attempted all eight questions and in that order, though a few candidates did Q1 (designing a questionnaire) and Q4 (calculations of sample size) first, and a few did questions in some other order. Both Q1 and Q4 were done well and performance of candidates on the paper as a whole was good.

Question 1

There were many excellent attempts at Q1 with most candidates drafting good questions on the topics requested. Some of the covering letters were longer than was necessary and a few candidates asked questions on extra topics. No marks were deducted for this, but the time spent might have been used better on answering other questions in the paper. A common error was to ask for the year of joining the organisation and then to request day and month as well as year. Another error was to ask about changes to the frequency of the newsletter and to offer monthly, the current situation, as a choice. Some of the rating scale questions only just qualified as being on rating scales.

Question 2

In Q2, a few candidates discussed the advantages and disadvantages of only one of the two methods. A few interpreted the one method as inserting a questionnaire in the newsletter every month, rather than just once as would be more sensible, but this interpretation was allowed. Not all candidates made clear which were advantages and which were disadvantages and some did not separate the two methods in their discussion, which made for muddled answers. Few mentioned that, with a random sample, it would be relatively easy to follow up non-response, but almost impossible when questionnaires were sent out with the newsletter. Some thought it would be necessary to draw up a sampling frame if sending the questionnaire to a random sample of members, although the frame clearly exists if newsletters are sent to all members.

Question 3

Q3 was not done very well. There were some very muddled answers with some strange ideas, particularly regarding taking a systematic sample from a list in alphabetical order, where some candidates were very concerned that people having the same surname would be chosen (unlikely unless the sampling interval were very small) and whether people with surnames starting with unusual letters would be chosen. For most purposes people's surnames are not connected with survey topics. In the case of some candidates, it was not clear that they appreciated how systematic sampling works.

Question 4

Nearly all candidates did the calculations in Q4 correctly, but not everyone made clear that they had checked that the sample sizes they suggested did not total more than the 95 for which money was available. A few candidates who noticed that a total was more than 95 then left the sample sizes unadjusted. In (iii), some did not answer in terms of why the method of (i) might be preferable to that of (ii).

Question 5

Q5 was done reasonably well, but there were also some muddled ideas. In saying that a pilot survey is a trial run, many candidates failed to say that it is usually on a small scale. Responses about a census were almost entirely in terms of large population censuses. Yet a census can be done on institutions and on small finite populations as well. In (ii) many candidates wrote about only one reason why results from a census might differ from true population values, whereas there are several reasons why there could be differences. Several candidates referred to a "census survey", which is something of a contradiction in terms.

Question 6

In Q6, it was not always clear whether a candidate knew what cluster sampling and stratified sampling mean, and the number of potential advantages and disadvantages of the methods mentioned were disappointingly few. As in other questions, some candidates did not make clear which was an advantage and which a disadvantage, and there were some muddled ideas. Some candidates described selecting a sample of **customers** although the question was about a sample of **supermarkets**.

Question 7

In Q7, most candidates mentioned that it was important to select customers of different types to interview, though some suggested what would be a biased selection by saying, for example, that customers who looked as if they were in a hurry should not be approached. Rather few candidates mentioned that it was important to select at different times of the day and week, and many failed to state where in the stores they thought the selection should be done. On the whole, comments on the conduct of the interview were better than on the selection of interviewees, though some candidates appeared to have missed the fact that the interviewers were provided with a questionnaire and that it was important to keep to the wording of the questionnaire, suggesting for example that interviewers did not use complicated questions. Hardly anyone mentioned the interviewer's appearance.

Question 8

In Q8, quite a large number of candidates suggested looking at till receipts or at loyalty card records, but these would not be observational studies that involve observing people's behaviour. Some candidates did not suggest using a checklist and so thought that recording and coding would be difficult. Although the advantages and disadvantages of the method compared with a questionnaire were not always clearly stated, and some candidates wrote as much about using a questionnaire as using an observational study, most candidates made some good points.

Ordinary Certificate Paper II

General

It was gratifying to find that almost all candidates had clearly listed the numbers of the questions they had attempted and secured any loose sheets in the examination book. It is better if graph paper is attached close to the question to which it applies. Almost all had started each question on a new page, as requested. If a question is answered in more than one place, candidates should make sure this is noted in both places. There were several charts and tables to be constructed. On the whole, the charts were better drawn than the tables. Each chart should have a title, axes and units clearly labelled, and any straight lines drawn with a ruler. Each table should have a title, with rows, columns and units clearly labelled. Separate the rows and columns by ruled lines where appropriate. Although it is acceptable for chart plotting to be done in pencil, the main body of questions should always be answered in ink. It is never appropriate to use red ink.

There is always some time pressure on candidates so it is important that they should read the questions carefully and only do what is requested: for example, overall % reduction omitted in Q3, wrong sectors considered in Q3 and comments omitted in Q6. Similarly, they should be familiar with short-cut methods of computation, particularly for the standard deviation and centred moving averages.

Question 1

This question was well answered on the whole, about one third of the candidates obtaining full marks. The main problems, apart from careless slips, were that some candidates did not know that average speed is calculated as the ratio of total distance to total time and others had problems knowing whether they were working in minutes or hours for the time calculations.

Question 2

Those who knew what they were doing found this an easy question. Several candidates, however, were unsure what is meant by a *contingency table* and simply gave the marginal totals. It would have helped accuracy if candidates had constructed a tally table first.

Almost all correctly stated the modal grade in Mathematics but some missed out the modal grade in English.

There were two methods of approach to finding the probabilities. The first was the relative frequency approach, simply working out the ratio of the numbers in each appropriate category. Others tried a formulaic approach, which was successful if clearly thought through, but some tried to apply formulae mechanically, often assuming independence or not sure which conditional probability was appropriate.

Question 3

Again this question was reasonably well attempted. Most of the bar charts were well drawn and included title, axis labels and source. A very few did not use graph paper or did not draw a bar chart. There was no need to rearrange the sources in order of contribution or to colour the bars, but this just wasted time, not marks. Most candidates could correctly work out the original pie-chart angles. When the percentage reduction in each sector was 10%, some candidates, even among those who realised that this would not affect the angles, forgot to say what the overall percentage reduction would be. Fewer candidates managed the last part successfully. A common mistake, through not reading the question correctly, was to assume that because Non-energy and Industry were the two reducing emissions by 20%, it was those two whose new angles should be calculated.

Question 4

The calculation of means and standard deviations should be routine for candidates and they should be able to demonstrate that they understand the principles of these calculations even if in everyday work they use a calculator to obtain the values. This question was not well answered on the whole.

In part (i), the Easy puzzles, the mean was 9 minutes and there were only 4 puzzles so it was easy to calculate the standard deviation using the formula containing $\Sigma(x-\overline{x})^2$. In part (ii), the Mild puzzles, there were 16 of them and the same times kept recurring, so it was appropriate to arrange the data in a frequency table. The mean was 11.75 minutes so in this instance it was preferable to use the rearranged formula for the standard deviation using $\sum fx^2$ and $(\sum fx)^2$. Many candidates were unable to demonstrate that they knew any formula for the standard deviation and also failed to give the units.

The **coefficient of variation** seemed to be an unknown concept to many. In the comments, it was important to distinguish between the interpretation of the absolute variation, as measured by the standard deviation, which increased with the level of difficulty of the puzzles, and the relative variation, as measured by the coefficient of variation, which showed that the Easy and Fiendish variabilities were similar, once the difference in the means was taken into consideration.

Question 5

This question was more successfully answered than probability questions in previous years, perhaps because the idea was a familiar one. A few candidates did not appreciate the implications of the phrase "without replacement" and used the same probabilities for each draw. There was some trouble with the idea of conditional probability needed in parts (ii) and (iii). The correct method is not to include the probability of a red sock on the first draw in (ii) and red and blue socks on the first two draws in (iii) for, if these probabilities are included, they need to be factored out at the end to obtain the correct conditional probability. In part (ii), many candidates appreciated that RRR, RBB and RGG would give pairs but forgot to include one or more of RRB, RRG, RBR, and RGR. Only the very best candidates spotted that it was in fact easier to calculate the probability of not getting a pair and then to subtract this value from 1.

Question 6

As in previous years, this question was not answered very successfully. Very many candidates do not seem to appreciate the meaning of $\Sigma(x-\overline{x})^2$, etc, how they relate to the correlation coefficient and how they can be calculated from the given sums of squares and products without going back to the original data. [In two cases, these are just the calculations needed to work out a variance and so should be familiar, the third is a simple variation of this idea.] Candidates can often calculate the correlation coefficient successfully, using a formula that they have learned without understanding what the separate parts mean. This lack of knowledge was apparent in the last section also when candidates quite correctly used the formula (new mean) = 1.8(old mean) + 32 to convert the means but then attempted to use the same formula for the other conversions. Only the best candidates appreciated that *r* and *b* would not change because the same linear transformation was being applied in each case.

Question 7

The definitions of terms were not well stated. In defining seasonal variation, the word "seasonal" or "seasonality" should not be used. Cyclical variations are not on the syllabus, but the word "cyclical", which refers to long-term, not necessarily regular, variations about the trend, is not appropriate in the context of seasonal variations that are short-term regular fluctuations about the trend. Many candidates stated when it was appropriate to use a multiplicative model rather than stating what a multiplicative model actually is.

The computations were in general well done with only an isolated numerical slip. Unfortunately, several candidates lengthened their answers unnecessarily by finding the average seasonal variations, which had not been requested and for which they received no marks.

Of those who did make comments, most identified that the trend was fairly constant but that the size of the seasonal fluctuations in the first and third quarters was increasing.

Question 8

Candidates demonstrated little understanding of percentage change, with only a handful obtaining correct values for the index numbers. More than one produced an index number with a negative value. The concept of a chain-based index was not well understood.

In addition to the usual comments about charts, it is appropriate here for the horizontal axis to be placed at 0 on the vertical axis so as to distinguish positive and negative values clearly.

A monthly percentage increase of 1.5% implies that the index is 101.5 based on the previous month and **not** 150 as many wrote.

The chart showed that the percentage monthly change was decreasing but, as these changes were positive from February through to May, costs were still increasing. June is the only month in which energy costs decreased as compared to the previous month and even then the actual costs were very similar to the April level.

Candidates need to remember to ask "Percentage of what and when?" before putting pen to paper.

Higher Certificate ("traditional" format)

Higher Certificate Paper I – Statistical Theory

General

In this final year of HCI, a pleasing overall profile of marks was seen. Failures seem usually to be the result of weak mathematics, often together with a lack of understanding of basic concepts. It is emphasised that mathematical notation must be mastered so that candidates can deal with algebra (sums, integrals) and calculus fluently and accurately; the distinction between discrete and continuous random variables should be clear, and concepts such as conditional probability and independence should be well understood. There were few infringements of the rubric.

Question 1

This was an average question. Part (i) posed difficulties for some candidates, with some mistakes in applying counting techniques in part (i)(c) in particular (although the law of total probability could be used, given answers to earlier parts). Attempts to "explain" part (ii)(a) were of variable quality, and some weaker candidates failed to obtain the probability mass function in part (b). Most attempts at $E(X)$ were sound, apart from the occasional confusion with integration.

Question 2

This question was less popular, but those who tried it achieved high average marks. There were many good answers to part (i) but more inaccuracies in part (ii). Use of the Bernoulli trial variables T, D and H to find the overall mean and variance of X (= $T + D + H$) was rare.

Question 3

This was a fairly popular question with about average success. In part (i), most candidates correctly found the distribution of $X - Y$, but there were a few "wrong tail" errors. Answers to (ii)(a) were usually good; the combinatorial factor 6 was sometimes omitted from the calculation for part (b) and, curiously, part (c) dealing with the mean height was quite often omitted. In part (iii) there were several errors ($E(X^2) = 172$ etc). Part (iv)(a) caused trouble: many candidates seemed unaware that a **mixture** (as opposed to a **sum**) of Normal variables is not Normally distributed; a few good intuitive arguments for non-Normality (e.g. based on the "two-hump" distribution of widely separated components) were seen, but not the expected approach of showing that $E(Z)$ was not the median of Z.

Question 4

This was an unpopular question, and one that appeared to cause particular trouble for candidates. Problems here seemed often to derive from mathematical shortcomings, e.g. difficulty with summation notation in part (i), misinterpretation of "more than two errors" in part (ii)(b). Part (iii) required application of the result of part (i) and the use of conditional probability, which was not easy for weaker candidates.

Question 5

This moderately popular question was seldom well done: part (i) in particular elicited mathematical nonsense from some candidates. It should be noted that "Show that …" requires a derivation from the data of the question, so that in part (i)(a) starting with a formula for the cdf is not appropriate. Answers to parts (b) and (c) were usually adequate, but many candidates were unaware of the *coefficient of variation*. Part (ii) was generally well rehearsed, apart from answers showing sloppy use of summation notation, sign errors, etc. Answers to part (iii) were often omitted, or else weak. This may have been due to mention of the coefficient of variation, but ratio comparison of the standard deviation with the mean for the given sample would have been acceptable grounds for rejecting the exponential model.

Question 6

This theoretical and practical exercise on the geometric distribution was, like Q2, relatively unpopular, but those who tried it were rewarded with good average marks. In part (i), arguments based on P(more than X failures) were accepted as an alternative to summing the probabilities. Explanations for the form of the likelihood in part (ii) were of variable quality; most candidates attempted to maximise $\log L(p)$, but this task was sometimes compromised by poor mathematics and often the maximum was not confirmed. There were some good answers to part (iii), although confusion of estimate with parameter and weak manipulation compromised some attempts.

Question 7

This popular question on a two-parameter discrete bivariate distribution was, overall, answered satisfactorily. Most candidates found k correctly; however, very few realised the need to check all the probabilities to find the tightest limits on θ. Part (ii) was well done, apart from some inaccuracies in seeking Cov(X, Y). Part (iii) confused several candidates who had (correctly) found Cov(X, Y) = 0 and (wrongly) thought this implied independence, or who had wrongly found a nonzero covariance and must have thought 3 marks a generous allocation for "hence not independent". Part (iv) was well done by most who tried it.

Question 8

This exercise on simple linear regression proved to be average in both popularity and difficulty. The initial statement of the model and assumptions quite often omitted the independence of the errors, but was otherwise usually sound. Most candidates plotted the graph well, but the necessary formulae and calculations were problematic for some candidates, especially those who were confused by the coding of the data. A useful discipline is to check if numerical answers are "sensible" in the context of the question (and to review the formulae and calculations if they are not). Part (ii) was generally good, with the dangers of extrapolation being well understood. The final part (iii) was weak. Several answers used z instead of t, or t on 9 df rather than 8 df, and the standard error formula was often misinterpreted as "s" and followed by an extra wrong division by $\sqrt{n}$.

Higher Certificate Paper II – Statistical Methods

General

Overall, this paper was satisfactorily done, with pleasing evidence of general competence in the operation of the standard test procedures examined in the syllabus. , statements of the theoretical forms and assumptions of the underlying probability models, and their interpretation, and the intelligent discussion of the results, were generally carried out less well.

Question 1

This question on the unbalanced one-way analysis of variance was unpopular and attracted only just over half marks on average. Surprisingly few candidates managed to state fully and accurately the model and assumptions for this procedure, as asked in part (i). Likewise in part (ii), very few answers included correct (or indeed any) calculations of the estimated residuals (given by $y_{ij} - \overline{y}_{i.}$ for the ith row and so not requiring detailed ANOVA calculations), and so there were few sensible comments on how well the model assumptions appeared to hold. Most candidates had a fair idea of the ANOVA procedure, although calculations were sometimes garbled. Least significant difference tests were sometimes attempted, although not asked for. Part (iii) was disappointing, in that very few candidates recognised the possibility that a linear regression might efficiently summarise the progression of estimated treatment effects with increasing dose level.

Question 2

This question was not popular, possibly because candidates saw the high graphical content as time-consuming (although the "made-up" plots asked for in part (i) could be done very quickly). It was not generally well done. Candidates varied in their ability to explain or illustrate the differences between R and r (as defined in the question), particularly the idea of R as linearising any monotonic trend. Some calculations of r were good, but weaker candidates were likely to have found the formula challenging to implement correctly. There were several good answers for R and for the associated test.

Question 3

This popular question was generally moderately well done, although with a wide spread of good and weak attempts. Several statements of the Central Limit Theorem (CLT) were very vague, entirely lacking any mention of the asymptotic distribution of the sample mean in terms of the parent population parameters and the sample size. This vagueness was sometimes reflected in part (ii), where the confidence interval sometimes lacked a $\sqrt{n}$ or was based on $z_{0.1}$, and part (iii), where there was appreciable uncertainty about calculation of the sample quartiles. However, most answers showed some awareness of why the CLT might not serve well in this example. Part (iv) discriminated strongly between weaker and more able candidates: many of the former failed to manipulate a probability statement to make μ the subject. In the final part (v), overt comparison of the two intervals was common, but the last two marks, for relating the skewness of the sample data to preference for the exponential model, were rarely given.

Question 4

This was a moderately popular question, done overall with about average competence. Here, as in other questions, candidates often lost marks by vague or incomplete statements of assumptions – thus, the requirements of constant mean and variance were often omitted in (i) and likewise the requirements of symmetry and consistent shape of distribution in (ii). Along with some good answers to the paired samples t test, use of the two independent samples test was also sometimes seen. In part (ii), the Wilcoxon signed rank test was usually selected, and the test statistic correctly found. Note, however, that a two-tailed test is required (to match part (i)), so the 2½% table should be used ($S \leq 8$).

Question 5

This question was rather less popular, but generally candidates achieved good scores. Part (i) was found easy by most candidates, and part (ii) offered marks for simple calculations from a given result, followed by a plot of the acceptance probability as a function of p, the proportion defective. Many candidates, however, found part (iii) more challenging. Some weak answers used a binomial (100, 0.01) distribution, and by no means all attempts made use of the correct Poisson approximation.

Question 6

This popular question was done quite well overall, but with a wide variation of standard. In part (i), supporting calculations were usually sound, but the degrees of freedom for the F test were sometimes wrong (e.g. (7, 8) instead of (6, 7)) and there was some confusion over the significance level (correct procedure is to compare the ratio (max. variance/min. variance) to $F_{\alpha/2}$ where α is the desired level). In part (ii) there were some sloppy statements of the null hypothesis (in terms of sample means rather than population parameters) and several garbled versions of the formula for the test statistic.

Calculations for the parallel rank sum test were usually accurate, albeit not always with the correct use of the table. If the rank sum of the smaller sample (of New subjects) is used (64), for the alternative of interest this should be compared with the upper (not lower) critical value, which can be obtained from the RSS Table as 71 (see solutions, (1+2+…+7 = 28) + right-tail CV for U (= 43), hence = 71). Comments in the final part were generally adequate: with correct calculation the tests agree.

Question 7

This popular question on the χ^2 test of fit was the highest scoring of all, bearing witness to the generally good preparation of students on this topic. Occasional errors were miscalculation of the probability (and expected frequency) of "6 or more" and rounding of expected frequencies to integers, and a few answers tested against the lower tail of χ^2. Part (iii) was less successful: few candidates made the point that the two tests taken together showed that the geometric model was plausible, but not with the parameter originally suggested.

Question 8

There were few attempts at this question on a randomised block design, and scores averaged just under 50%. There were few good answers to part (i), dealing with the form of the model and its theory and interpretation, but part (ii) was more successful. Here, as elsewhere, candidates seemed happier when carrying out well-rehearsed routines of calculation than when stating theory or interpreting results.

Higher Certificate Paper III – Statistical Applications and Practice

General

Q1 and Q7 were the most popular questions and Q2 the least popular.

Question 1

Many candidates phrased the hypotheses in terms of $\overline{x}$ and $\overline{y}$ instead of μ_X and μ_Y.

A number of candidates overlooked the denominator in the formula for s^2.

There was a general tendency to use only one critical point in the test decision, instead of making some attempt to refine the significance of results.

Overall, candidates scored well on this question.

Question 2

There were very few attempts at this question. The most common failing was to attempt to answer (ii) via ANOVA rather than estimating the effects directly from the means.

Question 3

The algebra in this question was carried out well in general. There was a general failing to check that the turning point in the likelihood function was indeed a **maximum**.

The comments in (iv) were generally poor. Candidates did not highlight effectively where the model departures were, or describe effectively some further examination of the two model parameters. The most obvious point missed was the desirability of collecting more data.

Question 4

Many answers to this question lacked coherence, in that a logical path in exploring the data was often missing. In (ii), examination of the data as a 3×2 contingency table was not effective. Good answers needed to focus on patterns within subject or type of course. In other words, appropriately constructed 2×2 tables provided the most meaningful results. A common unhelpful comparison was based on MSc Finance v. MBA Management. A number of candidates attempted to answer the question using tests of differences in proportions, but many failed to do this correctly. Using the χ^2 test is easier. The final section on providing a summary of the results was not done well. Again, general cohesion and clarity was lacking.

Question 5

Answers to (i) were generally weak, focussing only on each member of the population being equally likely to be selected. The definition of *systematic sampling* often omitted the need to select the first individual at random. There was some confusion regarding sampling v. non-sampling errors. These were often given the wrong way round.

Question 6

The various parts of this question where commentary was required tended to be poorly attempted, and often indicated lack of understanding, rather than just poor expression. The calculation of the irregular component was usually on the right track, though sometimes with the omission of correcting to make the seasonal components sum to one. In spite of a clear instruction with regard to plotting the irregular component, the plots provided were often incorrect in some way, either points were misplaced or axes incorrect.

Question 7

This was a popular question and attempts were generally fairly good. The main weakness was in the last part, where most answers did not really contribute any statistical point. For example, the possible statistical effects of those students who do not have any employment were not discussed.

Question 8

In part (i), the point generally omitted in the definition of a contrast was the fact that the coefficients must sum to zero. There was a range of errors in part (ii). Candidates often did not use the sample variances given in the table and embarked on lengthy and often error-prone calculations using the complete data set. Again there was insufficient refinement of the significance of the test result. In part (iii) the variance used was generally incorrect. The ANOVA residual variance was used rather than the variance of the relevant contrast (which involved the residual variance). Part (iv) was in general not attempted, and this would be the natural consequence of candidates having used the wrong form of variance in (iii).

Higher Certificate (modular format)

Higher Certificate Module 1 (Data collection and interpretation)

General

Q4 was very popular, with all but one of the candidates attempting it. Q3 was next in popularity and almost the same number attempted Q1 as attempted Q2. In general candidates need to make sure that they answer the questions set, not variations of these.

Question 1

In Q1, one or two candidates appeared to be writing about **sampling** rather than about **sampling frames**. In part (i), very few stated what imperfections occur in sampling frames. In (ii)(a), some candidates devised schemes linking the two frames although the question asked for a discussion of each frame. Most candidates who attempted this question wrote as if a sampling frame of **addresses** is equivalent to a sampling frame of **households**, which it is not. Insufficient details were given for adapting the frames to obtaining samples of households with school children.

Question 2

In Q2(i), some candidates discussed how to choose **patients**, although the question was on the choice of **hospitals**. In (ii), few mentioned that hospital records could be electronic or as some form of hard copy. Many focused on the kinds of information that would be useful to the research organisation rather than what the organisation would need to know. In (iii), some questions drafted by some candidates were not on post-operative care as requested. The six questions were for inclusion in a longer questionnaire in which it might be supposed that questions on other topics would be asked.

Question 3

Q3 was done fairly well. In (ii), on the main problems in obtaining information about people's annual incomes, the point that payments are made at different intervals of time was mainly missed, for example salary might be paid weekly but interest annually.

Question 4

Q4 was also done fairly well. In (i)(a), some candidates quoted values from the output but did not explain the terms as requested. A comment that the arithmetic mean is an average is not in itself an explanation of what it is. In defining the median as the middle value, few candidates mentioned what happens when there is an even number of values. In general there was little on interpretation of the measures, for example that the arithmetic mean is a sharing out average. In (i)(b), many missed the fact that the mode of zero (watching no television after school) would not be seen in a histogram. In (c), few commented that the distributions were skewed to the right. In part (ii), many missed important points as regards what advice to give a researcher who wants to conduct a survey of school children about their use of drugs.

Higher Certificate Module 2 (Probability models)

General

The high failure rate for this module reflects inadequate mathematical preparation on the part of many candidates. It cannot be said too strongly that fluent, accurate and confident mathematics is essential to the study of probability models, and that an understanding of basic ideas such as discrete and continuous random variables, cumulative probability and independence will be enhanced by mastering the mathematics in terms of which these concepts are defined. Areas of particular difficulty are highlighted in the question analysis below.

Question 1

This was a popular question, but scores were generally low. Answers to part (i) were often surprisingly weak, confusion of probabilities with numbers of possibilities, choices with arrangements, orders with or without replacement all being seen. Frequently, a wrong "formula" is guessed at and applied blindly. Explanations of the hypergeometric distribution in part (ii) were seldom good, but calculation of the probabilities and $E(X)$ often provided welcome marks to candidates who struggled earlier in the question.

Question 2

All scripts featured this question, which was generally satisfactorily done, giving the highest average score of this paper. Apart from some very weak work, answers to part (i) were generally good. However, in part (ii), the calculation of the cdf often failed because of a cavalier approach (or total indifference) to limits of integration. Sign and other errors also led to several wrong answers for $P(|X| \leq ½)$. Results for part (iii) were mixed, with most attempts based on a correct strategy but sometimes hamstrung by weak mathematics. Several students seemed unaware that $E(X) = 0$ by symmetry.

Question 3

This question on the Normal distribution was highly unpopular and not well done. Some students failed to make progress in part (i) by dealing with the linear combination $3X - 4Y$, but most of those who did completed the answer correctly. There were some good answers to part (ii), but very few made significant headway with the binomial distribution and Normal approximation in the final part (iii).

Question 4

This question on the Poisson distribution was popular but had the lowest average score on the paper. Most attempts to part (i) succeeded in establishing the recurrence relation, but several answers failed to mention its use for efficiently calculating the probabilities. Answers to part (ii) varied considerably in the quality of mathematics used to achieve the given results. The convolution argument required in part (iii) was usually poorly done, many candidates appearing to be clearly out of their mathematical depth. Part (iv) involved a practical problem: the first part (a) revealed very weak understanding of the probability calculation required, which was often rendered as

$P(A$ fails once$)/[\,P(A$ fails once$) + P(B$ fails once$)]$

instead of

$P(A$ fails once, B does not fail$)/[P(A$ fails once, B does not$) + P(B$ fails once, A does not$)]$,

(which reduces to the fraction 2/(2 + 0.5)). Answers to (iv)(b) were seldom done in terms of the total breakdowns variable $W \sim$ Poisson(2.5): instead, the required event was (invariably wrongly) expressed in terms of combinations of varying numbers of A and B failures, leading after much effort to a wrong answer. There were very few good answers to the Normal approximation in part (iv)(c). Candidates who got this far often used the unnecessarily inaccurate values 1.64, 1.65 for $z_{0.05}$.

Higher Certificate Module 3 (Basic statistical methods)

General

Apart from a number of very low-scoring scripts, the general standard of answers was in line with broad expectation. Most candidates are relatively well rehearsed in carrying out standard tests and inferential procedures, but their ability to discuss data in relation to necessary assumptions, and to interpret critically the results of statistical analyses, is less sure. Although not of prime concern in this module, there is some evidence of mathematical sloppiness in, for example, framing confidence intervals and statements of hypotheses in terms of sample estimates rather than population parameters.

Question 1

This was a fairly popular question with good average scores. In part (i), the population mean was usually correct, but the population variance was often wrong. Part (ii) was disappointing in that many candidates ignored the fact that sampling was **with** replacement. Samples were often counted without regard to multiplicity, so for example (1, 4) and (4, 1) are separate realisations of a set of 25 possible samples of size 2 with replacement from the original population of 5 items. In some cases, the sample means were not calculated. In part (iii), some candidates appeared to be under the mistaken impression that Var($\overline{X}$) = (population variance)/5.

There were several reasonable attempts at part (iv) (often requiring "follow-through" from earlier wrong variance etc), but very few candidates drew attention to the skewness of the exact distribution of $\overline{X}$ as a reason why the Normal approximation is likely to understate $P(\overline{X} > 16.5)$.

Question 2

This was an unpopular question with an average score of just over 50%. In part (i), there were few fully correct statements of the Normal approximation to the distribution of the difference of proportions from two independent samples. Many candidates failed to use the combined-samples estimate $\dfrac{r_1 + r_2}{n_1 + n_2}$ of the common p when constructing z for testing the null hypothesis of zero difference. Few candidates noted that , as samples are large, a valid approximate test is given by replacing p by its combined-samples estimate. In part (ii), some candidates used $n_1 = 920$, $n_2 = 994$, while others used the χ^2 test for a 2×2 table (which is asymptotically equivalent to the Normal test but was not the method asked for). With the help of "follow-through", there were several fair attempts at part (iii). However, the interpretation of the confidence interval usually offered was that it did not contain 0 (so that the null hypothesis of no difference in colour-blindness rates would be rejected). Although plausible (especially on the very clear figures of this example), this argument is not strictly correct, as the confidence interval is based on the separately estimated proportions p_1 and p_2 whereas the hypothesis test would be based on the combined-samples estimate of the common proportion. The intended interpretation was the frequency understanding of a confidence interval, namely that in repeated sampling approximately 95% of intervals so calculated contain the true excess proportion of colour-blind males relative to the proportion of colour-blind females.

Question 3

This popular question on the t tests and Wilcoxon tests for two independent samples was generally satisfactorily done, with an average score of 60%. As noted in other "applied" papers, however, while candidates are often well drilled in the mechanics of carrying out a test, they are often sloppy in stating the necessary assumptions and/or the null and alternative hypotheses (which are sometimes framed in terms of sample values rather than parameters). Candidates also often failed to note that the t test addresses the difference in means whereas the Wilcoxon test is concerned with the difference in medians. In carrying out this latter test there was confusion over the significance level (at the 5%

level the 2½% CV should be used as the test is two-tailed). In the final part, most candidates used dot plots (which probably give the clearest information-preserving visual comparison) or stem-and-leaf diagrams to compare these small samples; although both sets of data are clearly skew, this feature and its implication for preferring the nonparametric test were often not mentioned.

Question 4

This question on Normal-theory confidence intervals and tests on means and variances was also very popular, with scores averaging about 55%. In part (i), the means and variances of the observed data were usually found correctly, but common errors in part (ii)(a) were using z instead of t (and so ignoring the degrees of freedom), omission of the $\sqrt{n}$ ($=\sqrt{10}$) divisor, and quoting the subject of the confidence interval as $\bar{x}$ rather than μ. The χ^2 interval for σ^2 was less well known, although some good answers were seen. Similar corresponding comments apply to the tests of hypotheses in part (iii).

Higher Certificate Module 4 (Linear models)

General

Candidates overall achieved a fair average performance on this module, except for some lower scores that may have been associated with a poor command of English. Candidates were notably more successful on Q3 and Q1 than on Q2 and Q4. I suspect that, whilst Q1 and Q3 were perceived as relatively straightforward, this reflects particular difficulties with the paired data points in Q2 and with the correlation formula in Q4, rather than any deeper contrasts in the subject matter.

Question 1

This popular question achieved a good average score on the paper, but overall a wide variation in standard was seen. Part (i)(a) was essentially standard "drill" and was usually satisfactory. However, some weaknesses were apparent in part (i)(b), such as use of z rather than t and omission of the factor $\sqrt{\frac{1}{n_A}+\frac{1}{n_C}}$ when calculating the standard error relevant to the required confidence interval. It should also be noted that the ANOVA residual mean square should be used for this purpose, rather than a less accurate estimate of variance based only on the A and C data. Interpretation of this interval was usually given in terms of rejecting the hypothesis of equality of means (as the CI does not include zero), rather than as a process such that, taken over a long run of independent trials, 95% of the intervals calculated will contain the true mean difference $\mu_A - \mu_C$ say, as intended. Generally sound statements and suggestions were offered as answers to part (ii).

Question 2

This question was popular but usually very badly done, due mainly to poor handling of the relevant regression formulae and calculations, together with widespread failure to engage effectively with parts (ii)(b) (testing slope) and particularly (ii)(c) (pure error estimate of variance). In part (i), the graph was usually good, but commentaries usually omitted to mention the reasonably constant scatter and the fact that (as is a common assumption for regression) the independent variable was apparently preset without error. Some candidates' calculations in (ii)(a) went badly wrong after being confused by the pairing of observations at 6 different temperatures into taking $n = 6$ rather than 12.

Question 3

This question on interpreting computer-generated analysis of bivariate regression was only moderately popular but achieved a very high average score, from which it follows that many generally good answers were given, particularly to parts (i) and (ii)(a). In part (ii)(b), interpretation was generally sensible but often incomplete: thus, partial t values and the residual mean square and its basis for quoted standard errors were sometimes omitted. Many answers correctly commented on the prediction of a negative price for a carpet of zero dimensions and the danger of extrapolation far from the range of the data observed.

Question 4

This moderately popular question on parametric and nonparametric correlation measures was not well done. In part (i)(a), as elsewhere on this paper, graphical work was usually good. The linear trend was generally identified as appropriate for the use of r, but some weaker candidates confused the situation with regression, and very few mentioned the fact that (arguably) both variables here are subject to error (i.e. we have a genuine bivariate distribution). However, many weaker candidates had trouble with the correlation formula in part (i)(b), and many wrong calculations were seen. Although RSS Table 8 is most suitable for directly testing r, candidates using the equivalent t test were not penalised. Surprisingly, as Spearman's R has been well understood elsewhere in the Higher Certificate, there were few good answers to part (ii).

Higher Certificate Module 5 (Further probability and inference)

General

This paper includes developments of the probability results and methods introduced in earlier modules, including bivariate distributions and generating functions, and also some of the concepts and methods of statistical inference. The general standard was good, especially considering this was the first time the paper had been set.

Question 1

A popular, very well answered question. In Parts (ii) to (iv), it can be difficult to decide whether to work with decimals or fractions; either is correct, of course, but in this case it is probably easier to work with fractions with a common denominator of 48. In Part (iv), remember to say that values other than 2 to 8 have zero probability.

Question 2

A popular and quite well answered question, though not many candidates knew the relationship between the probability generating function and the moment generating function.

Question 3

Not such a popular question, but with some very good answers. Remember in part (ii) that, in order to establish consistency, you need to mention both that the variance of the estimator tends to zero as the sample size tends to infinity and that the estimator is always unbiased.

Question 4

Not such a popular question, with no very good answers. In part (iii), candidates often did not know the large sample property of the maximum likelihood estimator.

Higher Certificate Module 6 (Further applications of statistics)

Question 1

There were a few very good answers to this question, but many candidates did not know how to complete an analysis after obtaining the ANOVA for the Latin square. The estimated variance of a single observation is the residual mean square, and means could be examined using t tests (with the residual degrees of freedom). An F test could also be performed but is not really needed when specific comparisons are asked for. In this case, confidence intervals are more useful than significance tests.

Questions 2 and 3

There were surprisingly few answers to these questions. In Q2(b), the model without either of the x-variables should be checked also. In Q3, there was an almost total absence of the sort of phrases found in standard computer output, which can be a useful guide to problems with the model or the data.

Question 4

This was the most popular question. Two points gave trouble. First, when a target is given it is that target that should be the basis for the control chart of means, **not** the mean of the present set of data. [The mean of a **large** amount of **historical** data, obtained when the process was working normally, could be used if no target was specified.] Second, the limits for a mean (or total) of three items have a factor $1/\sqrt{3}$ (or $\sqrt{3}$) in them.

Higher Certificate Module 7 (Time series and index numbers)
Higher Certificate Module 8 (Survey sampling and estimation)

The numbers of candidates for these modules were small, and it is not possible to give detailed reports without identifying individual answers.

Graduate Diploma

Graduate Diploma Paper – Statistical Theory and Methods I

General

Generally, a disappointingly large number of scripts showed weak mathematics (sign errors, sloppy algebra, sums and integrals without limits, omission of domains for pdfs and pmfs, etc). Several candidates managed good answers to one or two questions but otherwise only scrappy, weak fragments.

Question 1

Very badly done by most candidates. Poor recall of formulae and weak understanding severely limited the scores attainable by most candidates. In parts (i) and (ii), lack of attention to the domains of pdfs of order statistics (and, correspondingly, the limits of integrals) was widespread. In the final part (iii), the hint to use $R = V - U$ was very seldom followed up.

Question 2

This was the most successful question, although by no means the most popular. There were several very good answers, and most candidates had little difficulty with parts (i) and (ii). Parts (iii) and (iv) were found somewhat more challenging, however. Candidates should note that the fact that W has the same distribution as $5 - V$ does **not** justify setting $W = 5 - V$ in part (iv).

Question 3

This question was both unpopular and poorly done. The general result asked for in part (i) was usually omitted or else attempted badly. Part (ii), based on the familiar situations of tossing a fair coin and rolling a fair die, was also surprisingly weak, with mathematical errors being common.

Question 4

There were many attempts at this question, several of which were derisory, being based on wrong or illegal transition probability matrices. In part (i), clear explanations for the Markov property were rare. Several candidates who did obtain the correct matrix also managed correct algebra to solve for the equilibrium probabilities in part (ii)(b) and obtained high final scores. Most candidates commented sensibly, if incompletely, on the fact that the average discount levels by no means fully reflected the differences in claim rates $\lambda = 0.1$ and $\lambda = 0.2$.

Question 5

This question on simulation by the inverse cdf method was the most popular of all, but candidates averaged less than half marks. In part (i), mistakes of algebra, wrong limits of integration, etc, were common, and few candidates correctly found the formula for a Pareto deviate x in terms of the U(0, 1) deviate u and general λ and α. Most candidates were even less happy with the discrete uniform distribution of part (b) and clear, accurate mathematical arguments here were very rare.

Question 6

This question on continuous bivariate distributions and transformations was also fairly popular, but not well done. Part (a) was more successful, although many candidates omitted the natural starting point of stating the joint pdf of Y and Z, and several subsequently fudged the argument for the pdf of T by neglecting either to integrate by parts or refer to the integral of gamma(2). Part (b)(i) was generally straightforward, but there were no good answers to part (b)(ii), often because of failure to use the substitution $u = 2v - w$ competently.

Question 7

This was a popular question, and, apart from the rump of very weak candidates using sloppy algebra, most managed reasonable answers to parts (i), (ii) and (iii). In part (iv), there were very few attempts via iterated expectations. Direct summation was preferred, but this often ran into trouble with wrong limits after cancelling x with $x!$, y with $y!$, but not realising that 0/0! = 0. Here and in part (v), some candidates also appeared to think that $p + q = 1$, which is not the case here!

Question 8

This was another unpopular and very weak question. Part (i) in particular exposed poor knowledge and fluency in integration by substitution; the conditional distribution, involving the simple quotient formula but no integration, was more successful, and in part (ii) the symmetry between Y and Z was usually well understood. However, part (iii) was poor. Candidates were intended to use the formula for $\mathrm{Cov}(X - \lambda Y, Y)$, but instead often tried to find the product moment by integration, an attempt that was invariably doomed to failure. There were very few attempts at part (iv) and none was mathematically completely correct.

Graduate Diploma Paper – Statistical Theory and Methods II

General

This paper aims to test a range of statistical principles and methods, and their application in simple situations. It was good to see some excellent scripts and few very poor ones.

Question 1

This was a popular, generally well-answered question. Faults included: in part (i), working with the exponential distribution with mean $1/\theta$, instead of θ; in part (iv), inventing a random variable $\overline{X}$; also in part (iv), claiming that an unbiased estimator of θ was θ/n.

Question 2

This was a popular question, with a spread in the quality of answers. Most answers to part (i) were correct, but they were often too long; it is important to drop terms not involving the data quickly. There were few correct answers to part (iv); several candidates just made a bold (often incorrect) claim without trying to justify their answer.

Question 3

Surprisingly for such a standard question, there were very few attempts. In part (b), when forming the likelihood, there is no need to worry about multiplicative terms not involving the unknown parameters.

Question 4

This was a popular question. Most candidates made some progress, but only a few obtained very good marks. In part (a), many candidates gave over-long answers, describing examples of pivotal quantities while the question was about pivotal quantities in general. A number of candidates wrongly claimed that that a pivotal quantity was a statistic. In part (b)(iii), a number of candidates found the equal tail confidence interval rather than the shortest one.

Question 5

This was a very unpopular question, but with some good answers. Presumably many candidates did not revise the topic of Decision Theory; if so, they are limiting choice of questions on the paper.

Question 6

Like Q4, a popular question in which nearly all candidates made some progress, but few obtained very high marks Answers to part (a) were sometimes too long with candidates mentioning Normal approximations rather than considering the area under the posterior distribution equal to the confidence level. In part (b)(iii), it is necessary to find the expected value of $e^{-\lambda}$ with respect to the posterior distribution; some candidates found the expected value of λ with respect to the posterior distribution and then substituted this value into $e^{-\lambda}$.

Question 7

This was a popular, well-answered question. Once again, answers to (a) were sometimes too long, with candidates describing the sequential probability ratio test rather than sequential tests in general.

Question 8

This was attempted by just under half the candidates, usually with a fair degree of success.

Graduate Diploma Paper – Applied Statistics I

General

Candidates followed the exam rubric. The majority answered two questions well, demonstrating a good grasp of both the theory underpinning a method and also its application in the scenario described in the question. Candidates who did not perform well overall could not maintain this level of expertise across sufficient topics/questions. The syllabus is quite far-reaching and it is important to study all topics in the syllabus.

Some candidates lost marks by not answering a question as asked. The paper is quite long, and it is worth devoting some time to reading a question carefully before starting to answer it. Little credit can be given for a mere repetition of the question; some credit can be given for a reasoned answer to the question (even if it is not correct) whereas an answer to a question different from the one posed cannot get much credit.

Good scripts provided reasoning and justification for answers given, and marks were awarded for good reasoning even when the answer was wrong or sub-optimal. No marks can be awarded for an incorrect answer with no reasoning. It is therefore important to show as much working as possible.

Care should be taken in all answers. For example, although plotting a graph is not in itself a high-level skill, deciding what to plot and how to select and label axes may require some thought, and marks can be lost by a candidate's rushing into plotting a graph without much thought and planning.

Question 1

Definitions were generally sound, although derivations of ACFs for the two time series were not always error-free. Sometimes, important steps in the working were omitted.

The answers identifying the time series were better when candidates gave their reasoning (even when the answers were incorrect).

Question 2

Many candidates answered part (i) in the context of principal components analysis, which was not mentioned in the question.

Part (ii) was generally well answered. Better answers gave more details.

Part (iii) was generally well done. Where working was shown, few marks were lost by errors if the overall method was correct.

Question 3

Candidates were either very confident with (i) or showed very little confidence. Candidates need to be familiar with this sort of bookwork. A review of previous papers will illustrate the types of questions that can be asked.

However, it is also important to be mindful of the marks awarded for a question. However long the answer, this part of the question is only worth 4 marks.

In part (ii), candidates who drew a clear graph showing the relationship between log(proportion) and dose were able to describe the data, to note the influential observation early on in their answer and then to interpret the diagnostic statistics and suggest further analysis. Not all candidates included the error in their model in (b).

Often a question like this follows through a line of reasoning, and it is important to try to establish the theme of the question – in this case, regression diagnostics.

Question 4

Part (i) is largely bookwork, which is commonly examined in this paper. Good answers scored highly here, but some candidates struggled to replicate the theory precisely. When the question is bookwork it is important to be precise in answers.

In part (ii), most candidates had an idea of what was required in (a), but many proposed a model with too many parameters. Good answers described the variables and their coding. Answers to (b), which required the choice and justification and coding of factors and an interaction, were less convincing. The major problem was a lack of precise detail in answers.

Question 5

It is not uncommon for a client to ask a statistician to assist in an analysis that is not suitable for a set of data. For the data presented in this question, neither discriminant analysis nor logistic regression will perform very well, owing to the nature of the data. For these 2-dimensional data, a careful inspection of the graphs and summary statistics should reveal this. Candidates who were not able to spot the problems with the analyses scored low marks n this question.

Question 6

Very few candidates attempted this question, and the answers were weak. This seemed to be symptomatic of lack of confidence in the basic theory.

Question 7

This question was based on a real scenario. Researchers can rush into multiple regression without thinking enough about how variables are coded, about relationships between predictor variables (and the consequent problems in interpreting results) and the dangers of automatic model selection. Also, they may not understand the assumptions required for statistical models and omit to carry out sensible checks, while doing other less important or even redundant checks.

There is therefore no single theoretical theme underlying this question: it is a concatenation of issues related to multiple regression in a practical situation.

The better answers were directly related to the question and gave reasoned explanations.

Question 8

This was a popular question.

Several candidates confused the random selection of doctors and nurses with the factor "professional group". The possible presence of an interaction should be obvious from an inspection of the data, but even if this was not done, the question specifically asks for a check for a "statistical interaction". Candidates who did not include this interaction in their model lost marks.

Most candidates could complete the ANOVA table and derive the MS; some errors were made in the derivation of the test statistics (i.e. choice of denominator). Good answers gave a practical (as opposed to mathematical) interpretation of the results.

Graduate Diploma Paper – Applied Statistics II

General

The overall level of preparedness of candidates sitting this paper was much better than in previous years. The paper covers a range of topics. Q2 on fractional factorial experiments and Q4 on response surface methodology were least popular, and mostly attempted by the more able candidates. Similar numbers attempted the other questions. The average marks ranged from 10 to 12, although it was somewhat lower for Q7 on ratio estimation.

As is usually the case, candidates were much stronger on standard bookwork than interpreting results of analyses in the context of the data/practical application given; answers to more descriptive question parts were often lengthy, but rarely well-structured. Some candidates showed expertise in two or three areas but weaknesses in the others; others failed to attempt all parts of the question, limiting the marks attainable. In general, candidates were poor at remembering simple formulae, particularly in the sampling questions.

Question 1

Most candidates could do parts (i), (ii) and (iii) relating to the design for a randomised complete block experiment, and construct the analysis of variance. Use of significance tests for comparing pairs of treatment effects was not well done. Many seemed unsure about how to apply the method of least significance differences. The interpretation of results was very poor.

Answers to part (iv) and (v) were disappointing. Only one candidate mentioned that there was no "control" fertilizer to assess the effect of the new fertilisers. Some candidates omitted part (v). Others appeared confused about how to partition the treatment sum of squares in the situation given.

Question 2

Few candidates answered this question, possibly because they were unsure about the analysis of fractional factorial experiments.

Question 3

The first part of this question covered the model for a Latin square design, and use of transformations. Answers were poor. Many candidates wrote down a model for a completely randomised design or complete block design. Only one candidate gave examples of situations where the analysis of variance of the raw data might be invalid, e.g. count data, survival times, etc. Many simply stated non-constant variance or non-Normal data, without further considering situations where such data might occur. There was a poor understanding of how to select an appropriate transformation in practice.

Parts (i), (ii) and (iii) were well done, generally, although some candidates had difficulty explaining the calculation of residuals.

Question 4

Few candidates attempted this question on first order models in response surface experimentation. Nevertheless, there were some good marks.

Answers to parts (i) and (ii) were poor. Only one candidate realised that the design was 2 replicates of a 2^2 factorial design with four centre points. Part (ii) was a standard textbook question on testing for lack-of-fit for the first-order model, but candidates seemed unsure about how to use the replicated points to calculate an estimate of pure error, so found it hard to complete this part.

The other parts were well done, which was pleasing. There were some good plots showing the fitted model over the experimental region, and the direction of steepest descent.

Question 5

Most candidates could identify the *standard* symbols used in stratified random sampling, but there were a few mistakes with defining the stratum weights, W_h (= N_h/N), some candidates suggested n_h/n or n_h/N_h.

A lot of mistakes were made in part (ii). Some candidates misread part (a), and calculated point estimates for the **mean** number of children in the region, rather than for the **total** number. Others confused the formula in part (b), and used **Neyman** allocation to determine the sample sizes in each stratum under optimal allocation. As the cost of sampling varies in different strata, the correct formula is $n_h \propto \dfrac{N_h S_h / \sqrt{c_h}}{\Sigma\, (N_h S_h / \sqrt{c_h}\,)}$. Few attempted to calculate an estimate of the variance of the estimated number of children in the region under optimal and proportional allocation, although the formula was provided.

Question 6

Answers to part (a) on sampling frames were often lengthy, and not that well structured; some parts were omitted. Some candidates misread the question, and gave examples of sampling frames suitable for selecting households only, so lost marks.

There were some good answers to part (b) on questionnaire design, but none of the candidates mentioned collecting data on time spent at the current address; it should be noted that assuming all households had lived in the area for 12 months could potentially bias the outcome.

Part (c) on sample size was fairly well done. Some candidates attempted to calculate the sample size at each of p = 0.35 and p = 0.55, and take the largest of the two sample sizes. Although, it gave the correct answer, since p = 0.55 is not too far from p = 0.50, the correct approach would be to calculate the sample size at the value of p (within the given range) with the largest variance.

Question 7

This question on ratio estimators was not done well. Some candidates failed to make a connection between the high correlation between the number of banana pits and the total number of banana bunches (0.7737), and the use of ratio estimators. A lot of mistakes were made in part (ii), possibly because candidates did not know the formulae for estimating the total number of banana bunches, using the simple random sample mean and the ratio estimator. In calculating the variance of the ratio estimator, most candidates replaced the population mean $\bar{X}$, although known, by the sample mean $\bar{x}$.

Only a few candidates attempted parts (iii), (iv) and (v), but those who did provided some good suggestions, such as cluster sampling, as alternative sampling schemes.

Question 8

The demography question was very popular, and attracted some high marks.

The definition of the *total fertility rate* caused some confusion in part (i). All candidates defined it as the ratio of the number of births to women 15–49 during the year to the number of women 15–49 at mid-year, which defines the *general fertility rate*. The total fertility rate is the sum of the age-specific fertility rates.

Most candidates produced nicely drawn age pyramids in part (ii), although a few candidates failed to adjust the mid-year population for the 0–4 age group, in order to make it comparable with the other age groups. Not all candidates commented on China's serious gender imbalance, when discussing the main features of the pyramid, which was a little surprising.

There were some good suggestions in part (iv) in relation to other demographic analyses that could be performed to explore the age-sex structure in China, e.g. rural vs. urban.

Graduate Diploma: Options Paper

The numbers of candidates for most components of this paper are small, and it is not possible to give detailed reports without identifying individual answers.