ROYAL STATISTICAL SOCIETY EXAMINATIONS, 2004

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 aspects of examination technique which could be improved. As we noted in last year's report, it is disappointing to see candidates losing marks unnecessarily. We therefore refer candidates again to the advice given at the start of last year's report, which can be found on the Society's website.

Matters raised by several examiners this year include the following:

- Read any question you intend to answer slowly and carefully, and ensure you answer the question actually asked. There is 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.
- Make sure you understand the difference between the instructions *explain* and *define*. An *explanation* of some concept requires one or more sentences, in which the concept concerned is described, and (if appropriate) the purpose or use is 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.
- Take note of the marking scheme printed on the paper. It is a waste of your time writing a two-page description of some topic, if this can only be awarded 3 marks.
- Ensure that you include sufficient reasoning in your answers for the examiner 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.
- If a graph is required, draw this accurately, using graph paper; make sure you include a title and label the axes. If only a rough sketch is required, this can be done in your answer book.
- Follow the instructions on the front cover of the answer book. We do realise that candidates will not want to spend time during the examination reading the front cover, 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. For example, candidates are instructed to start each new question on a new page.

Ordinary Certificate Paper I

The Ordinary Certificate syllabus covers the essential ideas of statistics in practice. Paper I on the collection and compilation of data includes the key topics of data collection in the field: what data should be collected, from whom, and how data should be captured.

The overall standard of answers was disappointing. The standard of English was in general poor, from both home and overseas candidates, and at times it was necessary to 'read between the lines' in order to follow a candidate's meaning. Answers tended to be written in a rambling style.

Too many candidates did not answer the questions set. Some appeared to have noticed one or two words in the questions and written on the topic these suggested to them instead of focusing on what was asked. Others answered questions at greater length than was required, for example in questions 3 and 8. Although some of these answers were of high quality, candidates who spend time on writing a fuller answer than necessary might have insufficient time to answer other questions. In three questions (1, 5 and 8) candidates were asked about advantages and disadvantages of methods. Many candidates, although making valid points, did not make clear which were advantages and which were disadvantages.

Although most candidates appeared to be familiar with the topics on the syllabus, many appeared to have little idea about the practical aspects of surveys or to have considered the possibility that samples are sometimes taken of objects rather than people. Some candidates had misunderstandings as regards the different methods of taking samples.

Question 1

Some candidates wrote at length about the advantages and disadvantages of random sampling (not required) as well as about those of quota sampling. Some made points which were advantages or disadvantages of both methods, that is did not make any contrast. Many candidates seemed to be unaware of what is meant by quota sampling and/or random sampling. Some appeared to think that all interview surveys are obtained by quota sampling, others that quota sampling means stratified sampling, some even that random sampling means interviewers stopping people at random.

Many wrote that there is no non-response in quota sampling, meaning only that quota targets can be met. However, there can still be non-response, since people approached might refuse to participate. Many candidates wrote about the difficulty of finding a quota of red-haired giants, but a design setting such a quota would be a poor design.

Question 2

Some answers to part (i) were about sample frames in general, although the question specifically required discussion of the use of a list of addresses to obtain a sample of households. Many of the suggestions made for overcoming the problems would not be helpful to someone using a frame in this way; for example, some candidates wrote that an up to date list should be used, others in effect that a different list should be used, or that the list should be supplemented by (in the UK) reference to the population census (information about individuals is confidential) or to the register of electors (but this could be used in preference to a list of addresses). As it was a list for delivery of mail some candidates wrongly thought the survey would be self-completion.

In part (ii), many candidates wrote about matters such as how interviewers could take a random sample, how to train interviewers, and how to advise them to start the interview (for example, with questions about the household composition). Although valid suggestions, these were more to do with the design of the survey and good survey practice than with helping interviewers choose a representative sample (for example, by deciding to select a quota sample). In both parts of the question many candidates seemed to be unaware of the difference between address, house, household, dwelling and building.

Question 3

Some candidates gave full explanations justifying the database they had designed. However, this was not required. Some candidates lost marks because they gave too few fields for some of the customer attributes, particularly the name and address. It is usually better to have several fields rather than one long field. As the database was to contain data about medication with the dates that a customer started and finished taking each one, provision for keeping details of more than one was needed. Not all candidates did this.

Question 4

Some candidates did not pay sufficient attention to what information was required from employees; for example, they omitted questions which were needed, or designed questionnaires which did not make it clear that a combination of methods of travel could be used in a single journey to work and that information was needed about the costs of each of these methods. A few candidates asked more questions than needed and one or two designed questionnaires for interview surveys. Some of the suggestions for three other questions were so open-ended that the responses would not have been very useful for someone seeking to improve the experience of employees in their travel to work. There were also some excellent suggestions of questions.

Question 5

Some candidates appeared to think that a diary in a dietary study would be rather like an appointments diary, and perhaps even that an appointments diary would be used for keeping a record of foods and drinks consumed. Others appeared to think that a diary would be more of a discursive journal in nature. Many wrote as if a diary in a study of this sort would have no structure to it, although structured diaries are used in many studies where subjects are asked to keep records over a limited time period, for example of expenditure or of television viewing, as well as in dietary studies. Dietary studies are more general than studies of slimming diets or of diets followed for medical reasons. Some candidates made points about diaries which held also for any mode of collection, and others wrote as if information collected by a questionnaire had to involve an interviewer.

Question 6

Many candidates wrote in general about simple random samples, although the question was on the issues involved in advising what size of sample to take. Some of these candidates wrote about what they called the lottery method which appeared to be a physical method of taking a random sample, but in practice paper or computer-produced tables of random numbers would be used. Some candidates said that as large a sample as possible should be taken, but clearly did not understand that this might be larger than needed if there were little variability in the population. Even the few candidates who mentioned that the sample size would depend on how close the researcher would like the estimate to be to the true value, and with what level of confidence, or who realised that the sample size would depend in some way on the variability in the population, appeared to be unaware that several quantities are usually estimated in a survey, each one usually resulting in a different estimate of sample size, or that if estimates are needed for subgroups a somewhat larger sample size would be needed than if estimates are needed for the population only.

Question 7

Most of the candidates were able to calculate the sample sizes correctly in parts (i) and (ii), but a few left these as non-integer values, and a few gave sample sizes which totalled to far more than the 40 for which the budget was sufficient. Hardly any of the candidates had much idea of the relative advantages and disadvantages of the two methods.

Question 8

Many candidates lost marks because they did not consider the possibility that N/n might not be an integer in systematic sampling. Others made no suggestion as to how to obtain a sample of about n books in the case of cluster sampling. Some candidates did not know the difference between cluster sampling and stratified sampling. A large number of candidates made little or no reference to the application of the sampling methods. In fact for all three methods of sampling some discussion of the practicalities of sampling books was needed, for example how to obtain a random sample of books from a shelf, how to decide on the order of books when taking a systematic sample. Since the purpose of taking samples was to estimate the average age of books, discussion was also needed as to how good or otherwise the methods would be as regards estimating this average.

Ordinary Certificate Paper II

Candidates are reminded that each new question attempted should start on a new page although it is not necessary for each section of a question to be on a new page. If a candidate returns to a question later in the answer book, it is helpful if this is indicated clearly on the page where the earlier attempt was made.

Candidates are reminded to read the questions slowly and carefully and to consider the marks allotted to a question when answering it. In several of the questions, candidates lost marks or wasted time by not having read the question carefully enough (these points are referred to in the specific comments). Most questions test a basic understanding of definitions and techniques but there is often a final section where an interpretation or comment is required. As one would expect, if only 1 mark is allocated to such a section, a simple comment will suffice; but a larger number of marks would indicate that a more substantial discussion is required.

Question 1

This question was not well answered on the whole. It is disappointing that these fundamental terms are so poorly understood. The same examples seemed to appear in many answers as though they were being reproduced parrot-fashion without true understanding. It would be beneficial if candidates could learn to spell *qualitative* and *quantitative*. There was a misconception that a discrete variable could take only integer values. Also many opined that a bar chart displaying nominal data had to display the bars in order of frequency. Although one may wish to do this, the bars can be arranged in any order; it could be alphabetical order or in the same order as an accompanying table, for example.

Question 2

The stem and leaf diagrams were better drawn, in general, than when a similar question was last set. Most candidates arranged the leaves carefully in vertical lines without the use of separating commas. Very few, however, labelled the units of the stem and the leaves and, although everyone indicated which catalogue was being represented, very few indicated that it was shoe *prices* that were of interest. Only a handful of candidates drew an unordered diagram as a preliminary and this caused some to have problems when they realised they had omitted values.

The majority could find the median and quartiles of these ungrouped data sets correctly, although a few used the $\frac{1}{2}n$th smallest value rather than the $\frac{1}{2}(n+1)$th for the median. Students should be encouraged to distinguish clearly between the *positions* of the median and quartiles in the sample and the *values* of these statistics. The writing down of some of the answers was sloppy in this respect.

A sizeable minority drew schematic boxplots; that is, boxplots that were not to scale. A clear scale labelled as `Shoe price ($)` was needed and it was beneficial if the plots were drawn to the same scale on graph paper. Almost all candidates need more practice in the interpretation of boxplots. For example, many stated that a narrower box implied that fewer prices were in that range, whereas it implies a greater concentration of prices. Much more careful use of the words spread, range and majority is needed.

Question 3
Candidates still have problems giving the maximum value for data values. For a value quoted as 2.7, to 1 decimal place, the maximum possible value is not 2.74 or even 2.749; for the purposes of this question it is right to quote the maximum value as 2.75 (even if it should strictly be 2.74999999999999...). The value 3.0 caused particular problems, since some candidates did not remember to treat the 0 in the same way as the 7, 8 and 4 in the other values quoted: the range of values should therefore be 2.95 to 3.05. Several candidates did not read the question carefully, and did not produce the values in a table.

In part (ii), some candidates wrongly calculated the mean of the original values and then wrote down the minimum and maximum values of this, rather than calculating the mean of the minima and the mean of the maxima.

Part (iii) was designed to test candidates' knowledge of the fact that, since the coefficient of variation is a ratio of two values, the minimum will be obtained with the minimum numerator divided by the maximum denominator and the maximum with the maximum numerator and the minimum denominator.

Question 4
This question was not very well answered on the whole. Candidates were not able to draw up the required table in part (i) from the values given and, of those who did get the table correct, many could not calculate the mean.

In part (ii), some candidates muddled percentage and proportions and treated 0.06 as 0.06%. Many candidates made the calculations much more complicated than they needed to be.

Question 5
The majority of candidates are now taking much greater care when drawing time charts. The main features to include, besides accurate plotting of the points, are a heading, clearly labelled axes, an appropriate vertical scale with the broken-scale convention used, if desired, and a key or clear labels to distinguish the trend line from the data points. Axes should be ruled, as should any other lines needed, for example to join trend values.

The data points should have been plotted over the three-week period, although a substantial minority of candidates superimposed all three weeks of data on a 7-day scale or, even more confusingly, plotted all the Monday data (that is, for Weeks 1, 2, 3) before all the Tuesday data etc. Not surprisingly this resulted in these candidates not realising that a 7-day moving average was appropriate for finding the trend. As a 7-day moving average is taken over an odd period, there is no need for the moving averages to be centred.

The definition of seasonal variation is still not clearly given by many, with the word ‘seasonal’ often repeated in the definition and the word ‘cyclical’ used, when cyclical variations refer to long-term fluctuations reflecting the overall economic cycle rather than the regular short-term fluctuations that are termed seasonal.

Most who had drawn the trend correctly commented on its downward slope but only a few noted that this was a steady decline and almost linear.

Question 6
This question was quite well answered on the whole. Some candidates failed to realise that all the necessary sums and sums of squares and products were given and they wasted time recalculating these.

Part (i) was quite well done; most candidates are now using divisor $(n-1)$ when calculating a standard deviation for a sample.

In part (ii), some candidates are still omitting either a title or axis labels from their scatterplot. Some of the scales chosen were not very sensible and made plotting difficult. Very few candidates marked the double point in any special way. Some candidates forgot to comment on their plot.

Although many candidates scored highly on part (iii), there were many mistakes in other solutions. Some clearly could not remember the formulae for the regression line and quoted something like the correlation coefficient instead. A substantial number misread the sum of squared values of x as 994.

Many candidates failed to realise that to draw the regression line they need to find just two points on the line. Some did not make it clear which points they had used, and some used 4 or more points. To obtain the two points, candidates could calculate y for any two values of x, or they could use the mean and one other point. The most elegant solution was probably one using the mean and the point which has to be calculated in part (iv).

In part (iv), the question asked for *reasons* whether the prediction was likely to be a good one. Most candidates mentioned extrapolation but did not give a second reason. (We had some sympathy with the one or two candidates who said it wouldn’t be a good prediction because they could not remember the formulae correctly!)

Question 7
Candidates continue to find probability questions difficult. Many candidates gave themselves a great deal more work because they failed to read the sentence: ‘I do a cursory check in week 1’. They thus drew a complete probability tree and then used probabilities of 0.5 and 0.5 for week 1 (or used 0.3 and 0.7 or some other variant). Some candidates did not mark the conditional probabilities along the branches but marked the total probabilities at the end of each branch.

Those who drew the probability tree correctly tended to do quite well with the calculations in part (ii). Nobody spotted that the week 3 probabilities can be obtained directly from the week 2 probabilities.

In part (iii), several candidates correctly calculated the probability of thorough checks in weeks 2 and 4 but then forgot to divide this by the probability of a thorough check in week 4 to get the correct conditional probability.

Question 8
This was essentially an exercise in the interpretation of index numbers in everyday language. A substantial number of candidates did not attempt this question but it may be that they had run out of time. The interpretations were quite well done and written in sentences rather than bullet points. It was pleasing that many candidates now know to use the term ‘percentage points’ when talking about differences between the values of index numbers with a common base. Several candidates ignored the information given in the question that age differences had been taken into account and suggested that the rates were high in Sale Moor and St. Martins because of the numbers of elderly residents. A few candidates realised that it was important to identify why these two wards had such high rates and suggested that social conditions or pollution may be factors. Some did not appreciate that low mortality rates were good news for the mayor.

Higher Certificate Paper I – Statistical Theory

The aim of this paper is to test the ability of candidates to understand and interpret basic statistical theory and to apply and adapt it to simple practical situations. Six of the 59 candidates had marks discounted for sixth or seventh questions whilst having five questions which scored at least as well.

The general standard was fair. Overall 33 out of 59 (56%) reached the nominal pass mark. Nine candidates obtained distinction standard (75% or more), three of these scoring over 90%.

Several instances of poorly organised work were noted this year. For example, solutions to different parts of one question were sometimes written several pages apart.

Question 1
Rather more than half of the candidates attempted this question, with mixed success (average mark 9.8/20). In part (i)(a), the answer 0.1296 was common, ignoring the possibilities that all four opposed or were indifferent to the complex; whilst in part (i)(c) the combinatorial multiplier 12 was often omitted. Most answers identified the binomial set-up in part (ii), but the conditional probability argument in part (iii) was less well done. In part (iv), the two cases ‘younger member opposes, the three older members do not’ and ‘one older member opposes’ must be considered: few good solutions were seen.

Question 2
A popular question, with marks averaging 12.3/20 reflecting general competence in the binomial and approximating Poisson calculations asked for in parts (ii) and (iii). However, the quality of mathematical analysis offered in part (i) was frequently very disappointing: weak and inaccurate notation was common, and indeed parts (i)(a) and (i)(b) were often omitted; even when the mgf was correctly obtained in part (i)(a) it was often not used in (i)(b). In part (iii), the percentage errors were usually correctly calculated numerically, but most candidates ignored the minus sign in the error for the approximate (Poisson) probability that there is exactly 1 wrong calculation.

Question 3
This was a highly popular question and achieved the highest average score (12.6/20). Half the marks were available for relatively straightforward Normal probability calculations in parts (i) and (ii), and

these were generally well done. Candidates' explanations for the probability in part (ii) being smaller were seldom fully convincing, however: very few candidates argued clearly that the mean volume of juice in a sixpack has the same expectation as the volume of an individual carton but only 1/6 of the variance, so that less of the distribution of the mean volume of a sixpack lies below 1 litre or 1000 ml. Part (iii) is best solved by introducing a symbol, x say, for the new (lower) mean volume setting (or for the amount by which the old setting may be reduced), and then obtaining and solving a simple equation for x. Most of those who tried this approach were successful, and several went on to deduce that the more accurate machine would pay for itself if 40000 or more cartons were sold.

Question 4

This was a somewhat less popular question, attempted by just over half of the candidates, and easily the least well done with average marks of only 6.8/20. In part (i), the Normal approximation to the binomial was well known, but the confidence interval calculation was surprisingly poor, with many muddles between the number of hits and the proportion of hits and many inaccurate calculations (often based on the variance rather than the standard deviation of the sample proportion). A few candidates tried the 'quadratic' formula here, which is acceptable but not necessary at this level; however, their algebra skills were insufficient to bring success. Part (ii) was quite well done, but there were no valid attempts in part (iii) at back-transforming the interval for p to one for $1-(1-p)^2$, the fallacious argument of using $p = 1-(1-0.6)^2 = 0.84$ in the standard formula being universally preferred. There were a few correct answers to the final part, but some attempts were frustrated by weak algebra.

Question 5

This question was moderately popular, and candidates achieved an average score of 12.1/20. In part (i), the graph of the exponential pdf was generally satisfactory, apart from starting at $t = 1$ (not 0) in some cases. Weak algebra (indifference to the limits of integration) led to several wrong answers for $F(t)$, but most candidates were able to deduce part (i)(c) from the given answer to part (i)(b) and went on to find the probabilities of settling an invoice within 1 week, between 1 and 2 weeks, and after 2 weeks, as required in part (ii). However, few were able to construct the likelihood from this information, and the mathematical challenge of the rest of this part was too much for all but the most able. Only a few candidates provided worthwhile answers to part (iii); some insisted on rounding the expected frequencies to integers and only very rarely were sound comments made (for example, that the exponential model predicted too few in the second week but too many at other times, so that the model variance or spread was too high relative to the mean).

Question 6

This rather mathematical question was relatively unpopular, with attempts averaging 9.1/20. Most candidates recognised that the given distribution was continuous, but (as in Question 5) integration to find the distribution function (not now a given answer) was very weak, with a consequent lack of success in finding the median and quartiles in part (ii), the moments in part (iii) or the results asked for in part (iv). However, it is worth noting that a few mathematically able students did very well on this question, emphasising its very straightforward nature in statistical terms.

Question 7

This question on estimation for the uniform distribution was the least popular of all, with only 20 attempts and a low average of 8.6/20. Part (i) was well-rehearsed, but only a few students saw the relevance of the distribution function to part (ii) and were able to find it. There were surprisingly few correct deductions of the pdf of the sample maximum, and attempts at the mean and variance were, yet again, often frustrated by sloppy integration. Several candidates correctly found the method of moments estimator in part (iii), but very few deduced its variance from part (i) and the general formula for the variance of a mean.

Question 8

This data-based exercise in simple linear regression was the most popular question, attempted by all but 7 candidates with a satisfactory average score of 11.8/20. In part (i), the data were generally well plotted (apart from a few cases where both x and y were plotted against the essentially irrelevant 'task number'). The existence of a good linear trend was well identified, and many candidates noted the influential isolated point (40, 53) although only a few commented that it was unlikely to be problematic because it lay well in line with the rest of the data. Part (ii) was mainly well done apart from occasional inaccuracies in the use of quoted least squares formulae; obtaining the residual mean square was the least satisfactory feature here. As a result, answers to part (iii) were often based on the wrong residual variance, and some candidates tested the null hypothesis of zero (not unit) slope (although a few correct versions were seen). In the final part (iv) (when not omitted), some candidates wrongly interpreted the 'regression through the origin' model as requiring a reanalysis of the data with the addition of the observation (0, 0); most others stated the correct model but either misquoted the slope estimate as $\bar{y}/\bar{x}$ or used their answer from part (ii).

Higher Certificate Paper II – Statistical Methods

The main aim of the Statistical Methods paper is to examine the understanding of fundamental concepts of statistical analysis. This is primarily achieved by asking candidates to solve standard problems of estimation and hypothesis testing, with particular emphasis being placed upon assessing each candidate's ability to summarise and interpret the results obtained from statistical analyses. Additionally, candidates are asked to describe or explain, sometimes with examples, particular concepts or general methodological approaches and provide some descriptive analyses of data.

In two questions, candidates were asked to provide a graphical presentation of data. The presentations were generally quite good, although some candidates still do not use graph paper, label axes correctly or choose a scale that is both appropriate and makes effective use of the whole sheet of graph paper. There are, however, many candidates who do not fully understand the purpose of particular graphs.

Many candidates, particularly those who appeared weaker in general, made errors in recalling the formulae relevant to the questions, even when they could identify the correct general approach to the analysis. We cannot emphasise enough the importance of understanding the statistical methods thoroughly. When a method is really well understood, the relevant formulae will come to mind naturally: rather than simply trying to commit formulae to memory, candidates should concentrate on learning the underlying methodology.

In this paper, candidates who correctly recalled the relevant formulae were generally very good at performing calculations although, as in previous years, many did not provide enough detail on the method of calculation.

Three questions required candidates to obtain and use confidence intervals (in question 2, confidence intervals were required for both mean and standard deviation). In many cases, candidates did not address the question as required. Whilst it was unnecessary to derive the (confidence interval) formula from that for the test statistic, some candidates did so whilst others merely tried to use the test statistic to answer the question. The choice of upper and lower 2.5% points from the relevant sampling distribution was not consistently made by those attempting to produce 95% confidence intervals, with some candidates using 5% points instead. Additionally, a number of candidates quoted limits for confidence intervals including inappropriate values of the parameter being investigated (for example, a population proportion outside the range 0 to 1, a correlation coefficient outside the range -1 to 1 or a negative value for the population variance) without taking further action or adding further comment to that effect. Overall, candidates' knowledge and understanding of the relevant formulae, calculation and interpretation of confidence intervals was not good.

Most of the questions required candidates to perform one or more hypothesis tests. Candidates did not always clearly state their null and alternative hypotheses. Moreover, when a one- (or two-) sided alternative was (or appeared to be) chosen, the corresponding percentage points of the relevant sampling distribution were not always selected. Candidates are strongly encouraged to try to understand not just *when* but *why* the upper, lower or both tails of the distribution define the critical region, and how this relates to whether a one- or two-sided alternative hypothesis is being tested.

Candidates could generally recall and use an appropriate test statistic in conjunction with the critical region to make a decision as to whether to reject a null hypothesis. However, when it came to interpreting the decision in the context of the application, for example by writing a report for a non-technical audience, many candidates were much weaker, and some simply did not attempt to provide what was requested. Likewise, although many candidates could state assumptions when using a model or statistical test, this appeared to be a feat of memory rather than showing real understanding; few candidates could explain whether assumptions appeared reasonable in a particular application, and several made statements regarding an assumption of Normality when considering categorical data! Other weaknesses when performing tests of hypotheses were:

- choosing a paired sample test when the samples were independent (but of the same size);
- using 5% points rather than 2.5% points for a two-tailed test at a 5% significance level;
- using incorrect degrees of freedom, and failing to justify the choice of degrees of freedom.

Overall, when performing hypothesis tests, candidates should give more detail of each stage of the test process. For example, statements such as '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.

Question 1
In part (i), a moderate number of candidates were able to write down the model suitable for a randomised block design, although several specified a model with only one factor. Most candidates defined the variables satisfactorily and stated the assumption of independently and Normally distributed errors with zero mean and constant variance. Few students indicated the range of the indices for the treatments and blocks and no student satisfactorily expressed parameter constraints imposed on the treatment and block effects. It is important to cover all the variables, parameters and indices when

defining a model. Also, when the model includes linearly dependent parameters, solutions should indicate a set of parameter constraints imposed (for example, sum to zero) which should be consistent with any explanation of the parameters (for example, the deviation from the overall mean due to a particular treatment).

In part (ii), several candidates wrongly performed a one-way analysis of variance (ANOVA), but many correctly calculated the relevant sums of squares for a randomised block layout, formed the corresponding two-way ANOVA table and completed the F tests. This question provided a good illustration of the appropriate use of blocking, as the test for treatment effect failed to approach significance if a one-way ANOVA was applied but was significant at the 5% level once the block effects were included in the model. Very few candidates attempted a formal follow-up test to identify which differences between treatment means were significant. Likewise, those making any consideration of the nature of the experiment and response variable used were few in number: candidates should firstly consider the nature of the experiment when assessing the likely validity of model assumptions (even in cases, unlike this, where residual diagnostics are available).

Question 2

Overall, this question was not well answered by many candidates. Many candidates inappropriately performed hypothesis tests rather than, or in addition to, calculating confidence intervals and several calculated them for either the mean or standard deviation rather than for both. A common error amongst those who did try to obtain a 95% confidence interval for the standard deviation was to use the upper and lower 5% points of the χ^2 distribution, thus obtaining a 90% confidence interval.

Interpretation of the confidence intervals or tests was limited, particularly in relation to the standard deviation. The fact that 0.5 was not rejected as a value for the standard deviation, despite being towards the lower end of the confidence interval, was commonly taken as providing evidence that the standard deviation satisfied the supermarket's policy of 'only buying tomatoes from growers who can supply tomatoes that have ... a standard deviation of no more than 0.5 cm'. It is important to remember that failure to reject a null hypothesis does not provide direct evidence that the null hypothesis is true.

Only a few candidates attempted to write a short report suitable for the board of directors, and a number of those who did failed to write in a clear, non-technical manner, giving a concise but justified recommendation. An important point, illustrated by the width of both confidence intervals, is that the sample size was actually too small to have a realistic chance of making a positive recommendation.

Whilst many candidates had some understanding of the weaknesses in the suggestion of sampling 8 tomatoes from each of 2 plants, their answers did not address the problem in terms of attributing the components of variability to within- and between-plant sources, and the corresponding effect on analytic method and precision of the estimation process.

It is important to answer the question posed and to reflect on the target audience. For most questions, the calculations, statistical methodology and associated interpretation should be directed at the examiner, as a professional statistician. However, when a question requires a report to be written for an audience who could be expected to have limited statistical knowledge, candidates should limit the report to non-technical material, except where absolutely necessary, in which case further explanation of the technical material should be included.

Question 3

Overall, this question was well answered by a moderate number of candidates.

In part (i), there was a satisfactory understanding of the relative merits of parametric and non-parametric tests. A number of candidates were under the misapprehension that non-parametric tests are easier to apply than parametric ones. This will be highly dependent on the facilities available to perform the analysis, the form of parametric or non-parametric technique to be applied and the size of the data set. Most candidates noted that parametric tests were more powerful that their non-parametric counterparts but omitted to state that this is in cases where both approaches are valid.

Almost all candidates drew satisfactory stem and leaf diagrams and most could use these to correctly conclude that the data were highly skewed and hence that a t test would be invalid. However, a disappointingly large number of candidates believed the study design to be paired, without any evidence provided in the question to support their belief. The majority of candidates used a one-sided test without justification. A common error was to believe that, when using the lower of two rank-based test statistics, the test was 'significant' if the test statistic value exceeded the critical value.

In common with other cases where two-sided alternative hypotheses lead to use of both tails of the sampling distribution, the result of a test is 'significant' if the test statistic is more extreme than the critical value. For the lower tail, 'more extreme' corresponds to 'less than' the critical value. (When using tables for non-parametric tests, candidates should be careful to check whether the values shown are just *inside* or just *outside* the critical region. This will be made clear in the table header.)

Question 4

This question was well answered by many candidates.

Part (i). As in previous years, a large number of candidates labelled the vertical (y) axis as 'frequency' or 'number of companies' rather than 'frequency density'. Most such candidates also represented the frequency by the height of the bar rather than its area. A small number of candidates failed to use graph paper or did not label both axes. Some histograms were squeezed into a small part of the graph paper; candidates should choose each scale carefully so that they make efficient use of the graph paper without the scale being difficult to use or interpret.

Part (ii). The estimation of mean, mode and standard deviation was, on the whole, done well. Relatively few candidates had errors in their calculations, although some who did not show full working lost marks unnecessarily. When estimating the median, a moderate number of candidates could not determine which data item represented the median (several seemed to automatically choose the 50.5th percentile without rationale!): it should be the $\frac{1}{2}(n+1)$th smallest value. Candidates who identified the position of the data item correctly could estimate the median correctly from the grouped data.

Part (iii). Most candidates identified that only estimation was possible due to the data being grouped. Some mentioned the fact that it was necessary to use an estimate of each data value (the mid-point) to estimate the required statistics. Others did not cover this aspect, but were credited for mentioning that only an estimate could be obtained as the data were a random sample rather than the entire population.

Part (iv). This part was answered well by many candidates. A few, however, did not estimate p directly from the data (typically assuming that the yield distribution was Normal and using this, together with the mean and standard deviation estimated in part (ii) to estimate p: this approach was inappropriate due to the apparent skewness of the distribution) and others could not write down the standard error of a proportion.

Question 5

Most candidates answered part (i) well, although a few confused Type I and Type II errors. Also, a considerable number of candidates provided definitions, rather than explanations, of the terms. In a mathematical context, a *definition* is a short and precise statement, usually expressed in mathematical notation. By contrast, an *explanation* will require one or two sentences.

While part (ii) was also answered well by a number of candidates, some were unable to make a serious attempt at it, and others incorrectly used the distribution of the data rather than the (sampling) distribution of the sample mean.

It is important for candidates to focus on the question: several computed the probability of accepting rather than rejecting the batch in part (ii)(a). Part (ii)(b), surprisingly, caused more problems despite its similarity with part (ii)(a).

Question 6

Part (i) was not well answered. Many candidates did not make it clear when the Normal distribution should be used and when, instead, the t distribution. The Normal is appropriate when the underlying population variance is known or when (the sample size being large) the variance can be assumed to be estimated accurately from the data. The t distribution is appropriate when the population variance is unknown, *provided* the underlying distribution can be assumed to be Normal. Few candidates covered more than one use of either distribution and few candidates made any attempt to give an example to illustrate the answer.

In part (ii), most candidates correctly calculated the sample means and variances for the new and standard batteries. Some, however, stated that their values were the population (rather than the sample) quantities, and a few candidates calculated sample variances using n rather than $n - 1$ as the divisor.

Most candidates stated null and alternative hypotheses concerning μ_1 and μ_2. (Note that these appear as μ_S (for Standard) and μ_N (for New) in the solution on the Society's website.) Some used one-sided and some used two-sided alternative hypotheses. Whilst either was acceptable, many of those who decided on a one-tailed test wrongly chose $\mu_2 < \mu_1 + 600$ as the alternative. This gave no opportunity to assess whether $\mu_2 > \mu_1 + 600$, which was the requirement for adoption of the new batteries. When applying the formula for the t test, many candidates did not include the specified value of $\mu_2 - \mu_1$ under the null hypothesis, namely 600, in the numerator. Several candidates did not correctly identify the degrees of freedom for the t test (18) and a few applied a z test, despite not knowing the population variances or having large samples. A number of candidates did not make the correct decision, rejecting the null hypothesis if their test statistic value was greater than the relevant percentage point from the t distribution and not rejecting it otherwise (irrespective of any earlier error). Most candidates who stated the assumptions recognised that one needed to assume that each sample came from a Normal distribution, although some failed to state that equality of variances was a necessary assumption.

Question 7

Most candidates correctly answered part (i)(a). Although many candidates computed Spearman's rank correlation coefficient correctly from the data, as required in part (i)(b), several did not use the ranks and therefore obtained a value outside the feasible range (of -1 to 1). Despite this, no candidate noted that the value obtained was clearly wrong – which they would have been advised to do – and some tried to interpret the nonsensical value thus obtained! A number of candidates could not correctly write down the formula for Spearman's correlation coefficient, several of whom omitted the factor (which was 6) in the second term. Most candidates performed a test of significance of the correlation coefficient, but most used tables relating to Pearson's rather than Spearman's coefficient. Few candidates interpreted the magnitude of the correlation coefficient; a value of 0.67 is often interpreted as a moderate positive correlation.

In part (ii), a number of candidates incorrectly treated the males and females as independent samples and, more worryingly, none of those actually recognised that the values in the cells of the table represented *pairs* of males and females (and therefore that the total sample comprised 100 rather than 50). Most candidates who treated the data as paired provided a good solution. As always, whilst some candidates provided an interpretation of the results in relation to the study, others merely reported failure to reject the null hypothesis (of no association) or reported merely that there was 'no association between males and females'.

Question 8

Part (a) was generally answered quite well. Most candidates recognised that the appropriate distribution was uniform, although not everyone adequately explained how they obtained the expected values. Likewise, most candidates determined that the sampling distribution (under the null hypothesis) was χ^2 and that there were 7 degrees of freedom. Very few candidates recognised the limited power of the test, with a small number commenting on the lack of a single ordering of the groups precluding the appropriate use of the Kolmogorov-Smirnov test although several mentioned the use of a cumulative distribution function in this test (a related point).

Part (b) was well answered by a number of candidates. Others could not obtain the relevant expected values under the null hypothesis (of independence) and several used incorrect degrees of freedom or percentage points for the null hypothesis sampling distribution, although most recognised this as χ^2. Several candidates made transcription errors when copying down the table of observed frequencies.

Higher Certificate Paper III – Statistical Applications and Practice

The aim of the Statistical Applications and Practice syllabus is to develop skills in data analysis, using the theoretical concepts developed in the syllabuses for the Ordinary Certificate and Papers I and II of the Higher Certificate, to analyse real data sets and communicate the results comprehensively. The questions on the examination paper require candidates to select and carry out appropriate statistical procedures and to report the findings and conclusions clearly. Candidates are also expected to be able to interpret computer output from statistical packages. Detailed knowledge of specific packages is *not* required.

Questions 1, 2, 3, and 6 were very popular, with almost everyone attempting question 6. Questions 4 and 7 were attempted by just under half the candidates, question 8 by just over a third, and question 5 by just under a quarter.

In three questions where a t test involving a pooled estimator could be done (not always the best method), the sizes of the two samples happened to be equal so the correct value of the t statistic could be obtained by finding the standard error as $\sqrt{s_1^2/n_1 + s_2^2/n_2}$. Candidates who are aware of this and use this expression should make clear that a pooled estimator of the variance is intended. In all cases the assumptions that the variances of the two populations are equal and that they both have Normal distributions should be stated.

Question 1

Part (i) was not done well. Many candidates gave details of the formulae for the test statistics (which were not required, and for which no marks could therefore be given), but did not state the null hypothesis under test (which *was* required). Further, a decision as to which test to perform must take into account whether the assumptions of possible tests are likely to be satisfied, so an answer to this part of the question did require assunptions to be stated.

Some candidates gave very confused or wrong answers to this part; for example, they worded hypotheses in terms of sample quantities, appeared to think that t tests were for small samples only, or that non-parametric tests were only for non-Normal populations. Many candidates who chose to do a t test in part (ii) seemed to be unaware that a pooled estimator of the variance was needed.

Question 2

In part (i), candidates should have realised that the collector of the data is not likely to know what assumptions it is reasonable to make about the populations of measurements and that there is therefore no point in asking about this. In part (ii), statements of the model needed to include definitions of the terms in the model. Some candidates did not state the assumptions well, and very few could make suggestions about how to investigate whether the assumptions were satisfied. Part (iii) was done fairly well. In part (iv), the residual mean square found in the analysis of variance should have been used. Some candidates used an independent samples t test (not the best method here) but did not make clear, or did not realise, that a pooled estimator of the variance was needed.

Question 3

A large number of candidates treated the samples as independent instead of paired in part (ii), even if they had referred to the pairing in part (i). In part (iii), many seemed to be unaware of the assumption that the variances of the two populations were equal, and did not make clear that a pooled estimator of the variance was needed.

Question 4

Most of the candidates who made a serious attempt at this question did it well. In part (iv), candidates were required to correct the year 2000 sales figures, but some worked on a different set of figures. Explanations as to whether the 2001 figures could be similarly corrected were not always right.

Question 5

Candidates perhaps did not think in enough depth about how to measure the seriousness of accidents and changes over time. Candidates need to have practice in doing open-ended questions like this involving discussion of data if they are to perform well on them in an examination.

Question 6
This was well done for the most part. Most candidates calculated the missing estimated coefficients from the data in part (ii) whereas they could be found more quickly from the output by multiplying the value of T by the SE of the coefficient. Many candidates lost marks on part (iv), omitting parts or giving wrong interpretations or explanations.

Question 7
In part (ii), free-hand boxplots received little credit, especially if it was obvious that there had been no attempt to draw them to scale. Graph paper should have been used. With one or two exceptions, candidates appeared to have no idea of any rule for defining outliers. Some showed one or two outliers on their diagrams but made no comment as to how they had decided on these values.

Question 8
Hardly anybody attempted parts (ii) and (iii). Some who did found wrong sample proportions.

Graduate Diploma Paper: Statistical Theory And Methods I

This paper examines probability theory – Bayes' Theorem, discrete and continuous random variables, univariate and bivariate distributions, transformations of random variables, simulation, order statistics, simple stochastic processes.

All but one of the candidates found five questions they could attempt. There were fairly similar numbers of attempts at all the questions, with the (usual) exception of the question on Markov Chains, which was attempted by only one quarter of candidates.

As was the case last year, candidates again seemed more comfortable reproducing standard proofs than solving problems. Candidates should be encouraged to broaden their experience of attempting problems, since this paper is unlikely to be passed on standard proofs alone.

Question 1
This examined the topic of joint discrete distributions. Candidates were expected to obtain the probability function of $X + Y$, when X and Y are independent geometric random variables. About half the candidates attempted this question but, with one glowing exception, their attempts were poor. Several candidates tried to use a method based on moment generating functions, but this is not helpful since the moment generating function of the sum is not likely to be immediately recognisable.

Question 2
This question was also about discrete random variables, this time examining iterated expectation and variance. There were some very good attempts, but the general standard of answers was poor. Several candidates assumed this question was about the continuous uniform distribution, in spite of the very clear description of the discrete uniform distribution at the start of the question.

Question 3
Candidates had to obtain the expected value and variance of the gamma distribution and then apply these results in a bivariate context where both marginal distributions were of the gamma form. Almost every candidate attempted this question and the general standard of answers was very high. It should be noted, though, that a number of candidates wrote $\Gamma(\alpha) = (\alpha - 1)!$, though this result only makes sense in the special case where α is a positive integer.

Question 4
This question tested candidates' knowledge of bivariate transformations. It was attempted by half the candidates but it was very poorly done. Many candidates seemed unsure how to apply the Jacobian method; for example, several forgot to invert the transform before taking partial derivatives. Hardly anyone was able to identify correctly the joint range of the transformed variables.

Question 5
Almost every candidate attempted this question, which examined moment generating functions, though with mixed success. Part (i), a standard proof, was well answered, but part (ii) was rather poorly done.

Question 6
Candidates had to apply their knowledge of sampling in order to obtain and compare the properties of estimators from simple random and stratified random sampling. Half the candidates attempted this question and several of them produced excellent answers. However, several candidates scored virtually zero marks for part (i) because they assumed the result they were trying to prove.

Question 7
The question consisted of three examples on simulation. The majority of candidates attempted this question. Some of them encountered considerable difficulty with part (ii), a more extended problem.

Question 8
The Markov Chains question was well tackled by a small number of candidates.

Graduate Diploma Paper: Statistical Theory And Methods II

The paper aims to test understanding of a range of statistical principles and methods, and their application in simple situations.

The questions that dealt with classical parametric methods (1, 2, 3 and 4) were each answered by well over half the candidates, as was the question on Bayesian inference (7). Just under half the candidates attempted the question (5) on non-parametric inference. Questions 6 (decision theory) and 8 (the essay question) were attempted by very few candidates. Questions 1, 2, 4 and 7 attracted a number of good solutions.

All but two candidates answered exactly five questions. The remaining two answered four questions.

A considerable proportion of the mark range was used across candidates. Overall the examination technique of candidates seemed to be good. However, some candidates wasted time, unnecessarily deriving results that were given in the question.

Question 1
There were many good answers to part (i), though some candidates wasted time by deriving the mean and variance results that were given in the question. Those who attempted part (ii) mostly gave good answers. Part (iii) was frequently poorly answered, usually because the factorisation theorem was incorrectly applied. There were surprisingly few attempts at part (iv) given that many candidates had successfully dealt with the method of moments in part (i).

Question 2
There were many good answers to part (i). Some candidates who used the hint wrongly assumed that

the score function itself (rather than its expectation) is zero. Some candidates ignored the hint and tried to derive the result from first principles with mixed success. There were some good answers to part (ii) and part (iii), but basic manipulation errors were not uncommon. For those who reached part (iv) the answers were mostly satisfactory, though, disappointingly, some candidates did not use Normal tables correctly.

Question 3

Part (i) was largely well done, though a few candidates wrote down the Poisson probabilities incorrectly. There were few good answers to part (ii). Some candidates seemed rather shaky on finding the variance. There were some decent attempts at part (iii) but few candidates spotted the link with the binomial distribution. There were few serious attempts at part (iv).

Question 4

Overall this question was answered well. A few candidates wasted time by not spotting in part (i) that the likelihood ratio is a monotonic function of the sample mean. Many candidates dealt with the SPRT part of the question particularly confidently and competently, which was pleasing.

Question 5

Most attempts at part (i) had some merit but were often rather woolly. It is always a good idea to define the relevant quantity (here the Kolmogorov-Smirnov statistic) clearly, as a first step; writing down a formula with suitable explanation is far better than writing a few rather vague sentences. There were no convincing answers to part (ii); for example, the fact that the power of the Kolmogorov-Smirnov (K-S) test may be low relative to parametric or some other non-parametric tests was not mentioned. There were mixed answers to part (iii); several answers had merit but some candidates clearly had no idea what the K-S test is and chose to make up their own tests.

Question 6

Very few candidates attempted this question. Those who did attempt it got the definition of risk wrong, which rather undermined their attempts at the rest of the question.

Question 7

There were many very nice answers to all parts of this question. A few candidates mixed up variance and standard deviation in part (iii) and part (iv), but in the main the attempts were pleasing.

Question 8

There were few attempts at this question, none of them particularly well done. The need to structure the discussion was largely ignored and there was no real evidence shown of breadth of knowledge of the subject of the question. A more reflective and wide-ranging approach is needed for this type of question.

Graduate Diploma Paper: Applied Statistics I

This paper is designed to test candidates' understanding of theoretical models and their practical application to realistic statistical problems. A basic grasp of theory is therefore necessary, but not sufficient, to achieve a pass mark. Candidates must be able to describe the purpose of a method, state the associated assumptions, apply it, and interpret the results in a way that a non-statistician would understand. Replication of bookwork, out of the context of the scenario in the questions, gains very little credit. Most questions have a theme that runs through them, so that answers to later parts relate to answers to earlier parts; few candidates were able to link parts of the questions.

The main weaknesses were:

- inability to state theory clearly;
- lack of attention to detail when describing data – candidates need to state what may seem 'obvious' to them as it might not be obvious to a client;
- failing to give a direct answer to a specific question asked – candidates need to read the questions carefully and then answer them, not just to repeat bookwork that seems related to the general area.

Average marks for questions ranged from 4.9 (question 5) to 11.3 (question 2).

Question 1
An applied time series question. Descriptions of the data were generally poor, and not many candidates were able to interpret the correlogram. Some candidates answered part (vii) in great detail and at length. Since there were only 4 marks available, there was no need to be too precise in specifying possible models.

Question 2
Bookwork on basic theory.

Some answers were quite good, but more attention needs to be paid to detail. Candidates should ensure they know definitions and proofs, since, when bookwork is required, answers must be precise. There is no merit in writing down ideas and comments that are not directly related to the question being asked. Again, candidates should answer *all* parts of the questions.

Question 3
The first part is basic theory, requiring models and diagrams. Few candidates submitted diagrams and the models were often imprecise. Working for the forward selection was often incomplete and badly described. It is not sufficient just to write down the 'best' model – candidates need to show that they understand the method. Interpretation of the model was generally poor and not related to the first part of the question.

Question 4
This style of question has not been used recently. However, it is the sort of task that an applied statistician meets very frequently. Credit was given for any valid ideas. As is common in practice, the scenarios did not give enough information for one to be sure of an appropriate approach. Candidates were therefore required to state what else they would need to know and under which circumstances approaches would be appropriate. Answers either tended to be too dogmatic (without statement of assumptions being made) or were far too vague, comprising ideas that were little more than a restatement of information given in the question.

Question 5
Few candidates attempted this question. It is a relatively straightforward application of theory. Graphs were quite good, but the descriptions of the data were often poor. Candidates should describe all the main features. Few candidates were able to replicate the theory about the exponential family, some

confusing it with the exponential distribution. Selection of models was not done systematically, and answers about the confidence interval were very poor.

Question 6
This question is about interpretation of principal components. Few candidates were able to explain why the use of the covariance matrix might be useful, or to interpret the analysis. Nor could they replicate the basic theory.

Question 7
The first half of this question was reasonably well done apart from the relatively simple matter of the Euclidean distance matrix in part (iii)(a).

Description of the dendrograms was generally poor, suggesting that although candidates could describe the basic purpose of cluster analysis they had little grasp of the effects of using slightly different methods. Part (iii)(c) required candidates to give possible reasons for the differences, not merely a description of the dendrograms. This related back to the data given at the beginning of part (iii).

Question 8
The nested design was correctly identified by most candidates. Those who identified the model were able to produce the analysis of variance table. However, interpretation of results was not always correct, and model specifications were sometimes imprecise. Most candidates described the design, not the data in part (i), although most identified the outlier in the data. However, few commented on the possible effect of the outlier in part (iv).

Graduate Diploma Paper: Applied Statistics II

This paper aims to examine candidates' understanding of the fundamental concepts of designed experiments and sample surveys, and their ability to apply these concepts to the analysis of data.

All candidates followed the rubric, with most candidates attempting exactly 5 questions. As in previous years, some candidates continued to lose marks by not answering the question asked or omitting sections of the question entirely. In addition, marks were lost as a result of very short answers being given for parts carrying a high number of marks.

The general standard was marginally better than that in previous years. Overall 10 out of 24 candidates reached the nominal pass mark. Unfortunately, nine candidates gained fewer than 40 marks on this paper. Two candidates performed exceptionally well, scoring marks of 60% or more. The average mark was mostly below 10 for all questions (excepting demography with an average mark of 12).

As is usually the case, candidates were much stronger on standard bookwork than on explanation and interpretation of results. For example, constructing an analysis of variance table and analysing residuals were generally done much better than commenting on results and explaining statistical concepts in experimental design. Often, data analysis ended at the F test in an analysis of variance, without further analyses being performed to explore significant interactions or differences between treatments. As in previous years, candidates seemed less familiar with the construction and analysis of balanced incomplete block designs.

In sample survey questions, there was a tendency for candidates to generalise the question, and give standard textbook answers, rather than to study the data collected, and explain why one technique might be preferred to another *for that particular set of data*. Numerical parts of questions such as calculating estimates of population means and standard errors were better done that those parts which required an explanation of statistical concepts or theory or sample size calculations. Question 7 was the most unpopular question, being based on elementary theory of estimators for the population mean and total in cluster sampling, and requiring the ability to apply this methodology.

A surprisingly high number of candidates answered the demography question on construction and use of life tables, and gained high marks. The topic of life tables has been less well understood in previous years.

Question 1 (19 attempts)
The answers to this question on the design and analysis of a 2^3 factorial experiment were somewhat disappointing. Part (i) asked candidates to explain, with the aid of a sketch, why blocking was preferred to complete randomisation. Most candidates overlooked the words *for this experiment*, and gave a standard text book answer on blocking, without considering potential sources of variation in the greenhouse. Only a few candidates attempted to sketch the experimental layout.

Most candidates completed the analysis of variance (ANOVA) table in part (ii)(a) correctly. There were a few errors; some candidates assigned more than 1 degree of freedom to the AB, AC, BC, ABC interaction effects (note that each term has a single df), and some forgot to subtract the block component in calculating the residual sum of squares.

In part (ii), many candidates concluded data analysis after completing F tests (of main effects and interactions in the ANOVA table). To gain full marks, candidates were expected to explore the data further, using formal statistical significance testing to assess the differences between means in the two-way table for the AB interaction. In addition, candidates failing to include a report for the scientist in part (ii)(c) lost marks. Marks were also lost by not including all of the key points in the report, especially that the level of factor C did not have an effect nor influenced the other factors in any way.

Question 2 (18 attempts)
Part (i) asked candidates to construct a balanced incomplete block (BIB) design to compare 5 treatments in 10 blocks. Most candidates seemed unable to write down this design, though values for N, r and λ were correctly determined.

Most candidates had difficulty calculating adjusted treatment means in part (ii)(a). A common mistake was in calculating Q_i (that is, $kT_i - B^{(i)}$). It should be noted that k is the number of units per block, and not the number of treatments. A few candidates were confused, some questioning what the treatments should be adjusted for, some mistaking batches as the treatments. Also, some misinterpreted the question to mean calculation of the adjusted treatment sum of squares. In BIB designs, any one block contains only a subset of treatments (since $t > b$), so the treatment means must be adjusted (for blocks) to allow for this imbalance in the design. Although a number of candidates correctly calculated the analysis of variance in part (ii)(b), very few used the computer output provided, which gave the residual sum of squares after fitting terms for treatments, batches, and treatments plus batches.

Overall, candidates correctly calculated the standard error for a treatment difference, and the least significant difference at the 95% significance level. Candidates failed to provide a clear explanation of the results.

Question 3 (12 attempts)
Candidates produced very short answers to part (i) which carried 14 marks. A common mistake was to state a method for using the residuals to check for Normality or heterogeneity of variance, without describing how to construct or interpret the plots. Answers describing how the residuals were calculated were often vague, and did not provide a numerical example. One candidate did not read the question properly, and wasted time actually producing a Normal plot of the data. Another candidate attempted to calculate sums of squares, as if the question had asked for an analysis of variance table.

Although the question clearly stated that '*a one-way analysis of variance was conducted*', a number of candidates wrote down, in part (ii), a model including a term for 'batches' (which was at odds with their earlier answer describing the calculation of residuals). All candidates correctly stated the assumptions underlying the analysis of variance.

A few candidates provided nice answers (based on the binomial distribution) as to why the data in Question 3 might not satisfy all assumptions. Most candidates failed to examine the data closely, and to note that the variances differed across the four conditions (groups).

Question 4 (8 attempts)
It was pleasing to observe that more candidates than in previous years attempted this question on response surface methodology.

Most candidates correctly defined the terms *rotatability* and *orthogonality* but then seemed unable to verify that the given design was orthogonal. In part (i), most candidates did not read the question carefully, and therefore did not realise that they were being asked to sketch the experimental design (for a central composite design), not write down the design matrix (or part of it).

Very few candidates attempted part (iii) which required constructing the analysis of variance using the computer output provided, which gave the residual sum of squares after fitting terms for linear, linear and interaction, and quadratic components etc.

Question 5 (15 attempts)
Both part (i) and part (iv) asked candidates to comment on the merits of stratified sampling and optimal allocation for the survey described. Many candidates ignored '*in this survey*', and gave standard textbook answers, without examining the data to see why these techniques might be appropriate in such circumstances. What was expected was a clear statement that the density of caribou differed dramatically based on habitat, so dividing the area into sub-areas might produce a gain in precision in the estimates of the population size. Optimal allocation was preferred because the population consisted of large and small strata, and because the stratum variances were much greater for large than small strata.

Most candidates correctly calculated an estimate of the population total and its standard error in part (ii), though calculations were notably lengthy, and prone to error.

Very few candidates attempted to calculate the total sample size and stratum allocations in part (v). The formula for calculating the sample size was given in part (iii), but candidates seemed unable to obtain V; that is, $(d/1.96)^2$. An interesting and noteworthy feature of these data was the sample allocation n_h in stratum ($h=$)3. This exceeded N_3, so required the stratum allocations to be re-calculated with $n_3 = N_3$.

Question 6 (13 attempts)
Part (a) was fairly straightforward, and adequately done. The key point is that regions of UK were strata, whilst postcode sectors were primary sampling units, and households were secondary units. That is, we have an example of stratified cluster sampling. Candidates failing to suggest a sampling frame of households in a country of their choice in part (a)(ii) lost marks.

Part (b) asked candidates to comment on the strengths and weaknesses of a printed questionnaire (awarded 14 marks). Often, answers were too short, suggesting candidates had not studied the questionnaire in sufficient detail to identify and include all key points. For example, not many candidates commented on marital status, which was limited to 'married' or 'single' in this questionnaire.

Question 7 (7 attempts)
This question on the theory and application of cluster sampling was not popular.

Part (ii) was a standard textbook proof, deriving the expectation of the sample mean for the cluster totals, using simple random sampling theory. In some cases, although the expected value was correct, there was very little explanation as to how the value was derived. None of the candidates recognised $\bar{y}_{cl} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} m_i}$ as a ratio estimator. Ratio estimators are slightly biased, but the bias is small when the cluster totals y_i are independent of the cluster sizes m_i.

Part (iii)(a), which required candidates to calculate an estimate (and its standard error) for the population total, was mostly well done. Very few candidates attempted part (iii)(b), even though this was straightforward, based on simple random sampling and estimators for the population proportion.

Question 8 (18 attempts)
This question on the construction and use of life tables was well done, and attracted some of the highest marks on this paper. In previous years, life tables has been one of the least well understood topics in demography.

There were a few errors in the calculation of expected age at death; some candidates tried to calculate these as the sum over all age groups of number of deaths multiplied by age.

Graduate Diploma Option: Econometrics

Question B1
This question was attempted by seven of 16 candidates who submitted answers to Option B, and the scores were disappointing. Whilst obtaining the least squares estimate of β_2 was straightforward, candidates did not appreciate the need to take expectations (based on the correct model) over the dependent variable values to demonstrate the required omitted variable bias. Correct use of this bias to deduce the likely result of a test of the true value of β_2 was rare. Answers to the evaluative parts of the question were of varying quality: although sensible points were made, no answer achieved a satisfactorily comprehensive coverage.

Question B2
Five answers to this question were seen, of which one was of reasonable quality. Generally, only perfunctory explanations were offered as to why the given model might adequately represent the reaction of consumption to changes in income. The long- and short-run multipliers were not always

correctly identified, but most answers did highlight some of the deficiencies in OLS analysis of this form of model. Similarly, in part (ii), the Koyck model calculations were disappointing although the qualitative critiques submitted in part (iii) were somewhat better.

Question B3

This question was very poorly answered by the five candidates who tried it. No fully convincing mathematical analysis was offered to express the given Keynesian conditions on mpc and apc in terms of the parameters of the two functional forms proposed. Such attempts as were seen showed weaknesses in both differentiation and logarithmic manipulation. Because candidates could not clearly relate mpc and apc to the model parameters, the t tests asked for in part (ii) were seldom correct. When comparing the models in part (iii), candidates focused on R^2, and sometimes on the Durbin-Watson statistic, but ignored the sizes, signs and partial significance of the model coefficients as well as the information provided to permit comparison of residual sums of squares on the same scale.

Question B4

Thirteen of the 16 candidates in econometrics answered this question, achieving a satisfactory average mark of 12.4/20. No candidates elected to write on assessing forecasting methods, but there was reasonably balanced coverage of the other five topics.

The concept of instrumental variables (IVs) was generally well understood, but few candidates explicitly noted that efficient IVs could be found by regressing the endogenous variables on all the exogenous variables in an over-identified system. Autocorrelation was similarly well discussed, albeit with few references to the classical AR, MA and ARCH time series structures, or to the Lagrange Multiplier test for higher-order serial correlation.

The nature and likely consequences of heteroscedasticity were well described, and references made to the Goldfeld-Quandt and Breusch-Pagan tests. Candidates were generally aware of the remedial actions of weighting or transforming the data, but failed to note that if a linear predictive model for y_t is valid then a linear predictive model for (say) $\log(y_t)$ cannot be valid.

Candidates were able to explain the idea of dummy variables and usually provided convincing examples of their use. However, the gain in generality from extending least squares regression modelling to the analysis of (co)variance and more general models involving categorical data was not always made clear.

Cointegration was generally well defined, although its extension to seasonal data was not mentioned. References to unit roots and the Dickey-Fuller test were made, but the relationship to vector autoregression models was mentioned only rarely.

Graduate Diploma Option: Operational Research

All four candidates attempted the critical path question (Question C3) and most gave good answers; one candidate obtained full marks. The optimisation question (Question C1) was very poorly answered – only one candidate appeared to know what the dual of an LP is. The highest mark for this question was 5. Only one candidate attempted the simulation question (Question C4) – and did quite well. There were no attempts at the queueing question (Question C2).

Graduate Diploma Option: Biometry

Candidates for this Option should make sure they can develop relevant topics from the compulsory papers in the context of biometric applications. Answers which do not do this will generally gain very few marks.

(No useful comments can be made here without identifying individual answers.)

Graduate Diploma Option: Statistics for Economics
Graduate Diploma Option: Medical Statistics
Graduate Diploma Option: Statistics for Industry and Quality Improvement

The numbers of candidates for these components of the paper were very small, and it is therefore not possible to give detailed reports.