ROYAL STATISTICAL SOCIETY EXAMINATIONS, 2007

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 2007 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 2007 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 of course 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 <u>not</u> need to begin each <u>section</u> 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 (<u>not</u> in the side panel) the numbers of the questions attempted. (The question numbers should be written <u>in the order in which you answered the questions</u>. 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 Paper I

General

Nearly all candidates attempted all questions. Although many attempted the questions in order, many did not, perhaps starting with questions they thought they could do well on. Questions 1 and 5 were straightforward and between them this year accounted for 29 marks out of 100, so one strategy could have been to do these questions first. However many candidates appeared not to know very much about methods of sampling (examined in Questions 3, 4, and 5).

Some candidates appeared to have taken little notice of the mark allocations. These are a guide as to how much candidates are expected to write as answers. If a section of a question is worth only 2 or 3 marks then a relatively short answer is expected and it is a waste of examination time to write a long answer. Conversely if a section is worth 6 marks or more then a very short answer is unlikely to be sufficient.

Care should be taken in reading the questions and in answering the questions that have been set. Credit is usually given only for answers or parts of answers that are on the question. There is little point in writing more generally on topics than is requested. Some candidates might have lost some credit because they did not make clear how their answers related to the question.

Question 1

This was well done by nearly everyone and some questionnaires were of a very high standard. Nearly everyone included an introduction to the questionnaire, but many forgot to tell respondents what to do with completed questionnaires. Questions were to be asked on how satisfied customers were with three different aspects of the holiday. Some candidates interpreted this very loosely and in some cases their questions would not have measured "degree of satisfaction". A few candidates also included questions on topics that were not requested in the question paper. By doing so they might have been short of time to do justice to the other questions on the paper.

Question 2

This was poorly done for the most part, suggesting that candidates were not familiar with what is required in coding responses or in drawing up a coding scheme. Some candidates did not read the question carefully and discussed responses to a different open-ended question from that described.

Question 3

In describing how to take a quota sample, few candidates made any suggestions as to how the manager should find people to take part in the survey. Mostly the suggestions that were made were to approach people either when checking in or when checking out of the hotel, both of which are likely to result in high non-response. In describing how to take a systematic sample, almost no candidates mentioned that, in finding the sampling interval, N/n (where N is the number who have registered in the week and n is the required sample size) has to be rounded to an integer. Many candidates wrote that an advantage of quota sampling is that

there is no non-response. However, there might well be non-response in this method but it should be possible to achieve the desired sample sizes in quotas. Few candidates mentioned the obvious disadvantage of systematic sampling that it might be difficult to make contact with members of the selected sample. Some candidates did not make clear what they thought were advantages and what disadvantages.

Question 4

Very few candidates correctly identified the methods of sampling in part (i) and even those who did appeared not to know that for cluster sampling to work well the distribution of views of patients in every ward (cluster) would need to be similar to those in the whole population of patients, and for stratified sampling to work well the views of patients in a particular type of ward (stratum) would need to be similar to one another and to those of patients in all other wards of that type. Some candidates wrote in detail about such matters as the need to take random samples and the need to have access to records of which patients were in which wards, which are necessary for the sampling methods to work at all but would not ensure that they worked well. Some stressed that some patients would not be in a fit state to complete a questionnaire, but the question does not say when they would be asked to complete it. Completion could be after they have left the hospital.

Question 5

Nearly all candidates did part (i) correctly, but very few proceeded correctly in (ii) after calculating $N_i s_i \big/ \sqrt{c_i}$. Some candidates lost marks in one or both of (i) and (ii) because they did not make clear what were the required numbers of patients to sample. In (iii) hardly anyone commented on whether the SD figures in the table would hold for current patients, or that the costs were estimates, or that the optimality property might not hold for other variables. Neither were there comments that using a uniform sampling fraction is safe in that it achieves representative-ness across all variables.

Question 6

This was done fairly well by most candidates, but some candidates made points for or against telephone surveys that applied equally to face-to-face interviews. In (i) not all candidates made clear whether the points made were in favour of telephone interviews or of face-to-face interviews. A similar lack of clarity also occurred in (ii), and in this part some candidates wrote about why face to face interviews were preferable to telephone interviews rather than why telephone interviews might be less successful that face to face, that is they changed the emphasis of the question. A few candidates appeared to think that face-to-face interviews only occurred when quota sampling was used.

Question 7

Although most candidates were able to state many advantages and disadvantages of using a longitudinal survey of last year's graduates and of using sample surveys of three years of graduates to collect information about graduates' occupations one, five and ten years after graduation, a sizeable number of candidates did not always make clear which were the advantages and which the disadvantages, sometimes even writing sentences which contained both. There were also some vague statements as regards costs. Selecting one sample for a longitudinal survey is likely to be cheaper than selecting samples from three different years

(though the comparison does depend on the sizes of the samples), but a longitudinal survey could cost more as it is run over a period of ten years. Many candidates missed the point that with a longitudinal survey it would be about ten years before all results of interest would be obtained.

Question 8

Part (i) asked for an explanation of why non-response is a problem in social surveys. Most candidates wrote something about the fact that responses would not be representative of the population of interest, but few mentioned bias in this context. Hardly anyone mentioned that the sample size would be reduced, resulting in larger sampling variances, or the extra cost involved in trying to reduce non-response. Some candidates wrote on the reasons for non-response, or the effect of non-response on, for example, provision of social benefits, neither of which was asked for. Part (ii) was answered well by most candidates, but some concentrated on problems with a sampling frame of addresses, such as buildings having been demolished, although surveys could well be sent to named individuals. Hardly anyone made the obvious suggestion that following up non-responders might reduce non-response, regardless of the reason for that non-response.

Ordinary Certificate Paper II

General

The overall standard was disappointing compared with previous years. A substantial minority of candidates were not sufficiently well prepared for the examination. Many did not read the questions carefully enough and failed to carry out instructions, particularly with regard to rounding values to a specified number of significant figures or decimal places.

The presentation of graphs and tables was satisfactory on the whole with suitable labelling and headings and ruled lines used.

Candidates should be reminded that basic formulae must be committed to memory, particularly those for standard deviation. Candidates should be familiar both with the defining formulae for variance, standard deviation, correlation and regression coefficients and with the rearranged formulae which are often more appropriate for arithmetic purposes.

When making comments after doing calculations, candidates should learn to use the results of the calculations in these comments rather than commenting solely on the original data.

Candidates are again reminded that questions, but not parts of questions, should start on a new page. They should write the numbers of all questions attempted in the space provided on the front of the first of the exam booklets.

Question 1

Most of the candidates were able to give an example of a discrete and a continuous variable (though a few cited a category variable as a discrete variable). They did, however, have trouble describing exactly what was meant by a discrete and a continuous variable. It is not sufficient to say that a continuous variable is continuous.

Candidates seemed much less familiar with the idea of inter- and intra-subject variation and many did not attempt this part of the question. Of those who did attempt it, surprisingly few muddled the two terms. Some candidates did not read the question properly and did not give a single experiment where both types of variation could be found.

Question 2

The stem-and-leaf diagram was on the whole well done with complete labelling. Some candidates, however, did not produce an ordered stem-and-leaf diagram. Others made mistakes principally because they did not check that they ended up with 31 observations.

The median and quartiles were generally calculated correctly, though a minority of candidates did not refer back to the original data as instructed. A common mistake was, after realising that the 16th ordered observation was needed for the median, to assume that this was the observation on Day 16.

Many candidates did not spot the obvious pattern in the data; this should have been apparent to them when they drew the stem-and-leaf diagram.

Question 3

Candidates were, in the main, adequately prepared for this question. To try to ensure that the cumulative frequencies were plotted in their correct position at the upper class boundaries, the maximum possible class length was asked for in part (i). Many candidates ignored this request and incorrect plotting followed. A substantial minority drew frequency rather than cumulative frequency graphs. It is usual for the cumulative frequencies to be plotted on the vertical, rather than horizontal, scale. It is clearest if the class boundaries are marked on the horizontal axis. It is unacceptable simply to mark the classes 20–29, 30–39 etc with no scale.

The estimation of the median and quartiles was generally satisfactory but candidates are reminded that they should state the results, and units, clearly in the answer book and not expect the examiner to read the values from the graph. Those who were unable to do the estimation satisfactorily tried to work from e.g. 50% of the data range rather than 50% of the total frequency to find the median.

Question 4

This question, in general, was satisfactorily done. This is to be expected as the correct calculation of means and standard deviations should be routine. Candidates did not always use the most convenient formula for the standard deviation and they should be aware of the different formats available to make calculation easier. Several calculated the variance rather than the standard deviation. Many seemed unsure about the calculation of the coefficient of variation. Its value should be stated with an accompanying percentage sign. The request for the results to be given to 3 significant figures was often ignored. The presentation of the table was good on the whole, although a few candidates included workings in the table itself (which was not appropriate).

The interpretation was not well done. Candidates did not "engage" with the data and realise that high levels of suspended solids was worrying regardless of the level of the other possible

pollutants. Many went back to the original data table without reference to the calculations they had just done. The very high values of some of the coefficients of variation (CV) should have alerted candidates to problems with the data. Those who did realise that high CVs indicated great variability in the data thought that this automatically meant the measurements were wrong and advised retaking them. They could have advised checking that there had not been transcription errors but measurements cannot be retaken in this situation.

Some of the better answers queried what was meant by "high" and "low" levels in this context.

Question 5

As usual, the probability question proved to be difficult for many candidates although many of the better candidates obtained full marks. Candidates did not read the question carefully enough and either omitted the village shop totally or included an extra venue and invented conditional probabilities.

In part (i) they failed to realise that in Week 1 I go to the hypermarket; there are no probabilities attached to Week 1.

When drawing a probability tree, candidates should always check that the probabilities on arms coming out of a node add to 1.

In part (ii), candidates did not realise that the conditional probabilities remain the same; it is only the situation in Week 1 that has changed.

Question 6

Answers to this question were rather disappointing and candidates must have wasted a lot of time on unnecessary calculations.

In part (i), the scatter plot was well done on the whole though more care needs to be taken in plotting points. Conventionally the x-axis is the horizontal axis and the y-axis the vertical axis and it is probably easier to stick to this convention. There was no need to add 1 minute 28 seconds and 1 minute 53 seconds to the x and y values respectively. With data of this type it is very much easier to employ the same scale on both axes, and to choose a simple scale to minimise the chance of mistakes.

Part (ii) was not well done on the whole. Some candidates wrote down the x values and added them to calculate the mean even though $\sum x$ was given. Most did not realise that $\sum (x - \bar{x})^2$ could be calculated from the summations given and how the product-moment correlation coefficient was related to the three quantities asked for. Most of those who calculated it correctly did so by subtracting the mean from each x value and summing the squares of the results, using up valuable time. Many got the correlation coefficient correct and in doing so obtained the three requested values but did not realise they had done so.

The comments in part (iii) were generally reasonable.

In part (iv), candidates failed to realise that a checkpoint time of 1 minute 30 seconds corresponds to an x of 2 (1.3 was a common mistake) and they did not convert their y value back to a finishing time by adding 1 minute and 53 seconds.

For part (v), no further calculation was required. By far the easiest and most accurate point to use when plotting a regression line is the centre of gravity $(\bar{x}, \bar{y})$ and, in part (iv), the value for $x = 2$ should have been calculated.

Candidates calculated correlation coefficients that were greater than 1 with no comment, and plotted regression lines which did not go near any of the plotted points.

Question 7

The explanation of terms was poorly done. Although most candidates characterised the trend correctly as long-term movement, they did not qualify that with the word "underlying" which is crucial. Several referred to "sideways" trend rather than the preferred terminology of a "constant" trend. Similarly, seasonal variations are short-term but what characterises them is their regularity. Candidates should not use the word "'cyclical" to describe seasonal variations and should be aware that cyclical fluctuations are long-term periodic variations about the trend, in general reflecting the overall economic cycle. Although most candidates used an additive model in part (ii), the understanding of what it is was not well demonstrated. Most tried to explain when an additive model might be appropriate rather than what it actually is.

In part (ii), a few candidates re-calculated the trend values although they were given on the question paper. Several tried to average the trend values rather than subtracting these from the data values before averaging. In general, most candidates knew what they were aiming to do even if arithmetical errors crept in. Some kept in far too many decimal places and did not round the final answer appropriately. No-one stated the units of the seasonal variation.

Question 8

The answers were very disappointing. The question was the last on the paper and index numbers are often the final item studied on the syllabus, but it is very worrying how difficult it is for candidates to interpret percentages, which is what index numbers are. It must be stressed again that quoting the value of an index number without stating the current period and the base period is meaningless. When interpreting the value of an index, e.g. 305 for year 2007 based on year 2000, candidates must be very careful to say either that it has <u>increased</u> by 205% from 2000 to 2007 or that in 2007 it is 3.05 <u>times</u> its value in 2000. The wording of many answers was insufficiently precise.

Many candidates did not appear familiar with the concept of a price relative or price relative type index numbers and reverted to calculating aggregate indices. Even those who managed to calculate an unweighted price relative index were unable to compute a weighted price relative index, even though they could write down the general formula for one. There seemed to be confusion between the weights to be used in the index, which should be expenditure weights, and the weights of ingredients in the recipe. Again, several candidates calculated an aggregate index instead. In the case of base weighting, the numerical values of an expenditure weighted price relative index and a base-weighted aggregate (Laspeyres) index are identical but this is not true for current weighting. More familiarity with price relative

indices should be encouraged.

The good candidates commented that the cost of ingredients was not representative of the cost of living and that other items would need to be considered.

Higher Certificate Paper I – Statistical Theory

General

This paper was found rather more difficult than usual, with 60% of overseas candidates and 57% of UK candidates failing. There is a very wide range in the standard achieved, with disappointingly many poor scores but also several good or very good marks. There were a small number of rubric infringements (candidates answering 6 or 7 questions), presumably reflecting the difficulty of finding questions that could be answered well. Less "routine" questions and questions involving algebra tended to be answered badly, except by mathematically adept candidates. Common weaknesses included algebra skills and failure to provide on request relevant verbal information or interpretation in clear and accurate English. Numerical work was usually accurate.

Question 1

This was a popular question, which was not too badly done overall. Most marks were gained on parts (i) (except for a few "with replacement" answers) and (ii) on the basis of (usually accurate) enumeration of possible cases. Part (iii) was weaker: several candidates scored zero due to arbitrarily using binomial or Poisson distributions to calculate the moments.

Question 2

This question was slightly less popular than Question 1 and less well done, averaging well below half marks. There were surprisingly many wrong statements of Bayes' Theorem, and several garbled or wrong explanations of the '1/8' probability in part (i), often omitting the key point that Mrs Smith's maternal grandmother must be a carrier. Parts (ii)(a) and (ii)(b) were usually correct. A common fatal error in part (ii)(c) was to assume that the condition of part (ii)(b) continued to apply (leading to the trivial wrong answer of zero); there were no fully correct answers to this part.

Question 3

This question on maximum likelihood estimation of the Pareto distribution was less popular but moderately well done. Weak candidates had trouble with sum and product notation in part (i) and many omitted to confirm that the turning point found was a maximum. In part (ii), very few students attempted to derive the asymptotic variance by means of the Cramér-Rao inequality as intended: most solutions consisted in substituting the estimate for the parameter in the given normal distribution and constructing the required confidence interval from this by rule of thumb. Part (iii) was often well done. Attempts at part (iv) were slightly less successful, often due to failure to compute the correct sample variance from the data given.

Question 4

This unpopular question on a bivariate application of the Poisson and binomial distributions was found very difficult. Overall, only a few good answers were seen, although in many cases a reasonable explanation was given in part (i). Several attempts at parts (ii) and (iii) showed a lack of understanding of marginal and conditional distributions, and/or inadequate mathematical technique to apply the concepts. In the final part (iv), there were a few correct answers to part (a) but very few candidates indeed identified the correct B(6, 0.2) distribution to use in part (b).

Question 5

This question on the geometric distribution was extremely unpopular among UK candidates, from whom only derisory answers were seen, but was less unpopular with overseas students, four of whom answered it very well. In part (i), several continuous graphs were given for the probability mass function, although the probabilities at 1, 2, …, 5 were usually accurately calculated. There were several good answers to part (ii), but part (iii) involving two variables was often omitted. Rather than use symmetry, several candidates computed $P(X < Y)$ as $1 - P(X > Y)$, thereby implicitly taking $P(X = Y)$ as zero despite the evidence of part (iv).

Question 6

This popular question testing routine calculations on the Normal distribution was only moderately well done by UK candidates, averaging just over half marks, and rather worse done by overseas students, who averaged 2 to 3 marks less. Part (i)(a) was generally good, but part (i)(b) led to surprisingly many errors in finding the variance of a sum of 100 independent, identically distributed Normal variables: use of the sloppy shorthand "100T" rather than the accurate form "$T_1 + \ldots + T_{100}$" often led to trouble. In part (i)(c), very few candidates appeared to realise that weights > 23.5 kg would be rounded and recorded as 24kg or more. Part (ii) was disappointing, as very few candidates correctly found $\mathrm{Var}(H + C) = 7$, leading to $T \sim \mathrm{N}(91, 43)$, etc. A wide range of answers was given to part (iv), and credit was given for all reasonable points.

Question 7

This unpopular question on the binomial distribution was very badly done, with only one script exceeding 50%. Most attempts established the recurrence relation, but very few indeed gave a convincing derivation of the mode, or produced an example of a double mode distribution, or showed generally that any such double mode must straddle the mean. Part (b) was badly misunderstood, with very few candidates realising that they needed to compare the total probability of the sequences (success, success, success; success, success, failure; failure, success, success) for the two cases (A then B then A) and (B then A then B). Some candidates correctly found the mean and variance of the total scores in the two cases, but several others applied binomial formulae despite the fact that the probability of success varies from trial to trial. Many answers to this part were scrappy and incoherent.

Question 8

This popular question on simple linear regression was fairly well done, although surprisingly many candidates incorrectly identified the dependent and independent variables in part (i).

The data were constructed to give numerically simple calculations (which were often well done), although interchange of axes and use of wrong formulae resulted in several wrong fitted lines. Despite the simple numbers, there were several errors in computing the residual mean squared error. Answers to the final part (ii)(b) were often compromised by a wrong root mean squared error carried forward from earlier work and failure to state the critical value of t (which meant that it was sometimes unclear whether a one- or two-tailed test was in fact being done).

Higher Certificate Paper II – Statistical Methods

General

Overall, the percentage passing (57% – 59%) was well up to that in past years. While there were a gratifying number of good scripts, there was still a wide range in the standard achieved. Questions requiring use of more complicated formulae tended to be answered very well or very badly. Common weaknesses included choice of the wrong analytical technique and failure to provide on request relevant verbal information or interpretation in clear and accurate English. Numerical work was usually accurate, except when poorer candidates applied algebraic formulae (e.g. for ANOVA or correlation), in which case errors were relatively common. There were very few infringements of the rubric, but weak students sometimes labelled worked pages of their answer books with the wrong question number, so increasing the risk of errors in marking.

Question 1

This was a very unpopular question, not badly done by UK candidates but poor for overseas.
(i) Relationship of limiting distribution to population parameters was often omitted, as was reference to "n large" or "$n \to \infty$".
(ii) Relationship to sample size in the context of skewness was seldom mentioned; references to known sd were rare; references to approximate z test were rare.
(iii) Errors of n for $\sqrt{n}$ or variance for standard deviation were very common, partly due to use of the sloppy notation "$Y = 100X$" instead of "$Y = X_1 + \ldots + X_{100}$". Few candidates bothered to interpolate the Normal table.
(iv) Wrong distributions (binomial, Poisson) were occasionally used for T. A critical region was sometimes given for the mean score rather than the total, and (if for the total) was seldom rounded to the correct integer values.

Question 2

A popular question: frequency tables and histograms (parts (i) and (ii)) were generally well and accurately done. Some students used 5-year groups 11–15, 16–20 etc instead of the intended 10–.14, 15–19 etc. Most candidates dealt correctly with the 5-year age groups, noting that age is usually quoted in completed years (although few if any allowed for this feature when calculating the median from the raw data). There were very few good attempts at finding the median from the grouped frequency data.

Question 3

This question on one-way ANOVA was unpopular with, and less well done by, UK candidates, although overseas students usually dealt with it satisfactorily. Whilst a small number of candidates answered very well, beyond stating the model and assumptions most were unable to convey any clear understanding of the rationale of the technique. There were also many errors of calculation. In the final part, some candidates included the t value as a factor of the required standard error estimate.

Question 4

Among UK candidates, this question on randomised blocks ANOVA was slightly less unpopular, and marginally less badly done, than Question 3; as with Question 3, overseas students were not put off and usually did better. Again, answers tended to be very good or rather bad, and there were several poor statements of the model. There were many errors of calculation, presumably reflecting lack of mastery of algebraic notation.

Question 5

This paired data question was perceived to be of average difficulty and was moderately popular. Several candidates lost many marks by failing to recognise the paired set-up, but those who did generally differenced the paired data accurately (though in one case the signs of the differences were ignored). Several candidates did not clearly state the null and alternative hypotheses, and this may have led to confusion between one- and two-tailed alternatives when choosing the t value for a test at the 5% significance level. There were several errors in calculating the median difference, even from correct data.

Most candidates are aware that nonparametric tests may be useful when the data are non-Normal, but few realised that the Wilcoxon signed-rank test assumes that the distribution under test is symmetric.

Question 6

This question, testing scatter plots and correlation, was fairly popular, with several good answers. Most plots were accurate, with well-labelled axes and reasonably well-chosen scales. Most candidates commented on the general linear trend and identified two or three likely outliers. However, the conditions on the data for validity of the Pearson correlation (bivariate, on interval scales, approximately linearly related) were seldom well stated. Good candidates found the correlation correctly, but others ran into trouble, probably because of weak algebra skills. Some candidates carried out simple linear regression analysis, although this was not asked for. Most knew that the coefficient of determination was the square of the Pearson correlation.

Question 7

Relatively few candidates attempted this question, but (particularly in the UK) those who did were rewarded with high average scores. Generally, the two independent samples set-up was correctly identified and the rank sum test was carried out accurately. However, several candidates failed to realise that the two-sided alternative hypothesis meant that the 2½% table of critical values should be used to give a test at the 5% level overall. Few students noted that the rank sum test requires the assumption that the distributions have the same shape, although

several practical points relevant to the context were mentioned (e.g. journeys should be made at the same time of day and avoiding special conditions such as road works).

Question 8

This question on the chi-squared test and a confidence interval for the difference in two proportions was popular but not very well done. Most students correctly identified the chi-squared test and applied the formula (usually without Yates' correction), though a few made their results less accurate by rounding the expected frequencies to integers. Clear statements of the conclusion of the test were often omitted. Part (ii) was less well done: several students used wrong proportions (e.g. 133/288, 460/878) and errors in dealing with the standard error term in the confidence interval formula were common. The final part (iii) required only a simple argument showing that quadrupling of n_1 and n_2 in proportion would almost certainly approximately halve the width of the interval, but this was seldom forthcoming.

Higher Certificate Paper III – Statistical Applications and Practice

General

Questions 1, 2 and 3 were the most popular questions and Question 7 the least popular.

Question 1

There was a general failure to demonstrate the actual degree of significance of results. Many candidates were content to simply reject H_0 at 5% significance and not show how highly significant the result. Some candidates wasted considerable effort by direct calculation of sums of squares instead of simply deducing these from the ANOVA table. A number of candidates made a poor choice of graph. Mean rating against percentage whey is a much better choice that mean rating against additive/no additive.

Question 2

A dotplot is the best illustration here for comparing shape, averages and spread. Many candidates gave the less useful stem-and-leaf plot. Hypotheses were often poorly presented, often in terms of sample statistics rather than population parameters, e.g. in terms of $\bar{x}$ rather than μ. There was a general failure in (iii) to mention the fact that the hypotheses here would relate to the median rather than the mean. The descriptions of alternative tests in (iii) were often poor on clarity.

Question 3

This was the most popular question but there were few very good attempts. Answers were often poor with respect to explanation and interpretation. Confidence intervals in (ii) were often given using z rather than t critical values. Also, the standard errors of $\hat{\alpha}$ and $\hat{\beta}$ were often erroneously divided by $\sqrt{n}$. Interpretation of residuals was poor, demonstrating little clear understanding of what was being looked for, i.e. lack of pattern, constant variance, outliers.

Question 4

A number of candidates tried to perform an ANOVA rather than a χ^2 contingency table test. Again, there were few attempts to refine the level of significance of the result. Many answers did not include an attempt at part (b). Those attempting (b) often omitted to use a common estimated proportion in the expression for the variance of the difference in sample proportions.

Question 5

Most attempts at this question scored good marks. For some unexplained reason, candidates often expressed the exponential distribution here in terms of λ instead of the given μ, and some penalised themselves in so doing. Few candidates checked that the turning point for $\hat{\mu}$ was in fact a maximum. Also, many candidates used $\frac{\partial^2 \log L}{\partial \mu^2}$ to derive an expression for $\mathrm{Var}(\hat{\mu})$ instead of finding the exact expression directly from $\mathrm{Var}(\hat{\mu}) = \mathrm{Var}(\Sigma_i x_i / n)$.

Question 6

A number of candidates gave plots by quarters rather than as a time series. Descriptions of the model and the moving average formula were poor, lacking evidence of understanding. Models often gave T = trend, and in that included the "average" response. A better description would be $X + T$, rather than just T.

Question 7

This was the least popular question and attempts were generally poor. There was little appreciation shown of the fact that the underlying distribution should be uniform, and very poor description of how this hypothesis would be tested. Further, the very important aspect of the answer here, that it should be for a non-statistical person to understand, was markedly absent. Many answers seemed to implicitly accept the notion of lucky/unlucky numbers!

Question 8

Weaknesses in answers to this question were again in the area of communicating concepts and ideas. In both (i) and (ii) there was little attempt to relate sampling methods to the most important aspect, that is, the good representation of the population. Many candidates simply did not follow the instruction in the question when it came to how to select and use the appropriate random numbers.

Higher Certificate Module 1 (Data Collection and Interpretation)

General

Most candidates attempted three questions as instructed. Questions 1, 2 and 4 were popular, but only a third of the candidates tried Question 3.

Question 1

There were many excellent answers to this question. In part (ii) a few candidates wrote rather more than was justified for a section worth 2 marks out of 20, perhaps commenting on everything they thought worth commending rather than concentrating on the features particularly worth commending as the question stated. In part (iii) it was pleasing that so many candidates found so many of the things that needed improving in the draft questionnaire. A few candidates were over-concerned with the physical layout of the questionnaire rather than the ways in which (the wording of) questions could be improved.

Question 2

In part (iii), although there were many good suggestions as to how the response rate might have been increased even more, few candidates commented on the fact that no attempt had apparently been made to contact those who had refused to respond to earlier mailings and that doing so might have been advantageous.

Question 3

The facts that this was attempted by such a small number of candidates and that (a) was not done well by those who did attempt it suggest that candidates had not covered this section of the syllabus. In particular, candidates did not appear to be aware that a finite population correction factor is needed when estimating variances of samples selected by simple random sampling. Some candidates did not continue to (b) although this was unrelated to (a).

Question 4

There were many good attempts at this. In (iii) many candidates gave as standard deviation the square root of the biased estimator of the population variance, that is the square root of "the sum of squared deviations from the sample mean divided by the sample size". The square root of the unbiased estimator of the population variance, that is the square root of "the sum of squared deviations from the sample mean divided by one less than the sample size", is more acceptable as a standard deviation of a sample. In (iv), few candidates commented that there was little overlap between the in vitro and in vivo figures, with all of the latter being smaller than the latter with the exception of one outlier.

Higher Certificate Module 2 (Probability Models)

General

Common weaknesses in scripts were poor integration and poor reproduction (or blanket omission) of requested routine bookwork. There were very few infringements of the rubric.

Question 1

Weak candidates were generally unable to solve this fairly popular set-theory question from the information given, and so did very badly; but most others applied suitable methods and, in the absence of numerical errors, scored well.

Question 2

This less popular question, on the exponential distribution, was on average poorly done by overseas candidates, while UK students averaged just over half marks. Part (i) exposed poor integration skills for several candidates, and there were few fully convincing answers to the conditional probability argument required in part (ii). In the numerical part (iii), several students calculated the sample variance by merely dividing the given sum of squares either by the sample size or by (sample size − 1).

Question 3

This averagely popular question on the binomial distribution proved difficult for many students. There were few good derivations of the distribution in part (i), and surprisingly many candidates were confused about the circumstances in which it is usefully approximated by the Normal. In part (ii), the correct approximating Normal distribution was usually found, but (much to the detriment of accuracy) it was often applied without the continuity correction; not all candidates realised that the binomial table provided in the RSS book of tables is of cumulative probability. Several students did not bother to calculate the percentage error, and some of those who did failed to convey the direction of the error. In part (iv), very few students indeed attempted the derivation of the negative binomial probability; however, most of the calculations in this part were accurate.

Question 4

This question, on Bayes' Theorem applied to Poisson data, had the lowest average scores (well below 50%). Very few adequate arguments for Bayes' formula were seen in part (i), most candidates merely stating the result, sometimes wrongly, for at best partial credit. Despite this, the problem part (ii) was often well done, suggesting that the principle was understood better than its accurate mathematical enunciation. There were also several good answers to part (iii), but one student was tempted to regard two sample lengths each containing two flaws as equivalent to one length containing four.

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 Paper – Statistical Theory and Methods I

General

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. The distribution of overall marks this year was again bimodal – about one-third of the candidates made very good attempts at the paper, while the remainder were able to achieve very little. This pattern was repeated in many of the individual questions. Even standard proof work was tackled with mixed success. All candidates bar one found 5 questions that they could start, but some of these were token attempts. Candidates should generally work to improve their ability to explain what they are doing (e.g. in simulation) and should lay out their working more clearly.

Question 1

This examined bivariate discrete distributions, in the context of combining independent Poisson random variables. Every candidate attempted this question, though the mean mark they achieved was one of the lowest for any question on the paper. Few candidates were able to apply the results of parts (i) and (ii) to help solve part (iii).

Question 2

In this question, candidates were required to work with a bivariate hypergeometric distribution. They were expected to explain why such a distribution would arise in a particular context and why this meant that various marginal and conditional distributions would be hypergeometric. Three-quarters of candidates attempted this question, many of them quite successfully. Others appeared very confused about the circumstances in which hypergeometric distributions are appropriate models for random variables.

Question 3

This examined regression, in the probability sense. Only three candidates attempted this question but their mean mark was the highest for any question on the paper.

Question 4

This is a standard type of question examining bivariate transformations; this year, the context involved independent random variables that follow standard Normal distributions. Two-thirds of candidates attempted this question. Although they did fairly well in part (i), most struggled to establish the link between parts (ii) and (i) and, therefore, scored few marks for part (ii).

Question 5

Again, this was a standard type of question, about moment generating functions. Almost all candidates attempted this question, again with mixed success. Again, some candidates started the question well but found it hard to work their way through it, using the results of earlier parts in their later answers.

Question 6

Only one-third of candidates attempted this question on order statistics. There were two excellent attempts; the remainder were disappointing, though for a wide variety of reasons.

Question 7

This question examined the probability integral transform and simulation. About three-quarters of the candidates tackled this question, generally well. Surprisingly many of them had problems explaining why the probability integral transform could not be used in a context where the distribution function was complicated, and even more did not recognise that a gamma distribution with integer index can be simulated by a sum of independent exponentials.

Question 8

This question, on Markov chains, was attempted by fewer than one-third of candidates. This question had the lowest average mark of any question on the paper. There was only one reasonable attempt at it. Several candidates had clearly misunderstood the context, so that they were trying to model it by a chain with just two states.

Graduate Diploma Paper – Statistical Theory and Methods II

General

The paper aims to test understanding of a range of statistical principles and methods, and their application in simple situations. All but two candidates answered exactly five questions; the others answered fewer than five. There were two outstanding performances where the candidates scored high marks on every question they attempted. Some other candidates scored well, gathered marks efficiently and showed good examination technique. However, there was a disappointingly large proportion of candidates who appeared to be poorly prepared for the paper. For example, the likelihood function is an important ingredient of many of the methods and applications featured in the syllabus (e.g. sufficiency, maximum likelihood estimation, hypothesis testing, sequential methods, Bayesian methods), yet many candidates could not form a likelihood for observations from a standard distribution, could not calculate a log-likelihood from a likelihood, and could not differentiate a log-likelihood. Another area of common weakness was that several candidates appeared not to understand the difference between a parameter and an estimator.

Question 1

About half the candidates attempted this question. There were some good answers to all parts of the question. In part (a), some candidates wasted time by converting (not always successfully) a Normal random variable to standard Normal from first principles. In part (b), those who made more than token efforts made reasonable progress. A surprising area of difficulty was in turning an inequality for V into an inequality for θ in order to construct a confidence interval for θ.

Question 2

The majority of candidates attempted this question. Results were mixed, including some very nice answers. Parts (i) and (ii) were bookwork, but many candidates made no progress here and appeared to have no real understanding of the notion of sufficiency. Part (iii) was rather poorly done with few satisfactory attempts. Uniform(0, θ) was the most popular successful example. There were some good answers to (iv) but some candidates ignored the fact that the sample size was n; if you only have a sample of size one, then the notion of a sufficient statistic is rather trite.

Question 3

Almost all candidates attempted this question. Those who could construct and handle the likelihood tended to do well, whilst those who could not necessarily made poor progress. This was evident in parts (i) and (ii), but also some candidates did not use the hint and/or forgot to obtain the method of moments estimator. In (iii) a common error was to overcomplicate the form of the large sample confidence interval. There were some good attempts at (iv) but some candidates just considered the case with $n = 4$ without looking at the required more general case with $n \geq 4$.

Question 4

Few candidates attempted this question and there were no really good attempts. Only two candidates did a good sketch of the loss function for part (i) and there were no convincing answers to (ii). For (iii), the fact that the sum of two independent exponential random variables has the given gamma distribution was rarely noticed, which made calculations for part (iii) needlessly complicated. Those that got as far as part (iv) made reasonable progress with it.

Question 5

Most candidates attempted this question. There were some good attempts at parts (i) and (ii) but some candidates were unable correctly to find the likelihood for a sample from a Poisson distribution and others could not handle Normal probabilities. Part (iii) was mostly rather poorly done with many opting to use a generalised likelihood ratio test rather than deducing the uniformly most powerful test. Only a few candidates knew what a power curve looks like.

Question 6

The majority of candidates attempted this question, which had the highest average mark. Part (i) was bookwork and many candidates had a reasonable grasp of this. There were no good answers to part (ii), where many candidates treated expected sample size as if it meant actual sample size. In part (iii), the general results were put in the context of the SPRT for a Normal problem. Some candidates were unable correctly to find the likelihood for a sample from a Normal distribution. Those who got as far as part (iii)(b) answered it well.

Question 7

The majority of candidates attempted this question. Part (i) is essentially bookwork as it is a standard textbook question. It was designed to give a straightforward path into the more challenging part (ii). There were some nice answers here but once again some candidates

tripped up in the derivation of the likelihood, while others took a prior distribution on X rather than on μ, thereby missing a key aspect of Bayesian inference. Those who struggled with part (i) made little progress with part (ii). The sketch of the curve in (ii)(a) caused surprising difficulties. A few made it though to part (ii)(b) and there were some decent attempts, but once again the curve sketching was not impressive.

Question 8

About half the candidates attempted this question. Solutions were often lengthy but rarely well structured. Description of the two tests for part (a) was bookwork but usually poorly done. The issue of parameter estimation was not tackled with confidence. In part (b) the attempts were mainly poor. Some candidates wrongly suggested Spearman's rho as a nonparametric procedure for two independent samples. Most candidates correctly suggested a two-sample t test for the parametric procedure but few were able accurately to write down the test statistic or the degrees of freedom. Comparison of the two procedures usually lacked substance.

Graduate Diploma Paper – Applied Statistics I

General

There were some very high marks – from candidates who read the questions carefully, answered them as they were posed, and were able to interpret the results of analyses in the context of the data/practical application given. General definitions (e.g. of an interaction) will be awarded few marks if the question is asked in a particular context. Simple tasks like plotting graphs and describing correlations need to be done carefully and with attention to the context. The syllabus refers to interpretation of computer output, and the applied statistician is often asked to interpret output from a package with which she or he is unfamiliar. Candidates need to practise looking at computer output and reading it carefully; some of the answers to the questions are in the computer output. The paper covers a range of topics. Some candidates showed expertise in two or three areas but weaknesses in the others. The specific topics examined in a year are relatively unpredictable, and it is unwise to ignore topics. Candidates should study the models, their assumptions, how to interpret output and be prepared to discuss limitations of the methods.

Question 1 (Time Series Analysis)

This was generally well done, although there were examples of careless or incomplete derivations and limited understanding of the application of the methods.

Question 2 (Principal Component Analysis)

Good answers related the output to the variables described in the question. The major weakness related to the choice of number of components. This is not always trivial, and the simple criteria used by some packages (as here) should not be used uncritically.

Question 3 (Discriminant analysis/logistic regression)

Few candidates answered this question, possibly because they were unable to interpret the

output, or they were unsure about logistic regression. Several candidates interpreted the output as linear regression not logistic regression.

Question 4 (Generalised linear modelling)

Candidates were able to do the model selection, but less able to describe the theory behind the model. There is usually a question on generalised linear modelling, and it normally asks for some theory about the model. When preparing for the exam, candidates should make sure that they can write down models and derive standard results about them.

Question 5 (Backward elimination in multiple regression)

Most candidates could do the first two parts of the question, although working was sometimes messy and difficult to follow. Automatic methods for model selection (like backward elimination) are not without problems. The quote in part (iv) is genuine. Candidates should be more familiar with the problems (e.g. multiple testing, over-fitting) and show less confidence in automatic procedures.

Question 6 (Model selection in multiple regression)

Some candidates failed to identify the response variable (ozone concentration), and showed too much confidence in automatic methods of model selection (see Question 5). Model selection can be informed by knowledge about the variables; candidates should be more familiar with this and show an applied approach to data modelling.

Question 7 (Design matrix)

Few candidates attempted this question. Although the application is rather unusual it is a basic application of the design matrix in regression. Candidates found it difficult to apply the design matrix to the problem given and some showed little evidence that they understood the role of the design matrix.

Question 8 (Analysis of variance)

This is a nested model with random effects (apart from fixed treatment effects). Candidates who correctly identified the model and associated analysis did very well. If the analysis was not for a nested model then few marks could be awarded. If you are going to answer a question like this you need to be really sure that you have the correct model.

Graduate Diploma Paper – Applied Statistics II

General

This paper aims to test candidates' understanding of fundamental concepts on designed experiments and sample surveys, and their ability to apply these ideas to data. All but one candidate answered exactly five questions. Several candidates ran out of time on the final question (fourth/fifth), but tended to score some of the highest marks among candidates on individual questions. The overall level of preparedness of the candidates sitting this paper appeared much better than in previous years. There were some good performances, but most

candidates still got less than half marks. Two candidates performed exceptionally well, scoring marks of 60% or more. Almost all candidates showed at least some understanding of the material. Questions 1, 3, 6 and 8 were popular questions; Questions 4, on response surface design, and 7, on cluster sampling, were least popular with candidates. The average mark ranged from 8 to 10, although it was somewhat lower for Question 5 on stratified sampling with proportions. As is usually the case, candidates were much stronger on standard bookwork than on explanation and interpretation of results. Often, data analysis ended at the F test in an analysis of variance, without further analyses being attempted to explore significant interactions or differences between treatments. In general, candidates were poor at considering the practicalities of sampling.

Question 1

In part (i) the explanations given were often disappointing. The field was bounded by a river and motorway, in directions at right angles. A Latin square arrangement was used to remove any systematic variation from these two known sources. Many candidates failed to mention this.

Part (ii) was reasonably well done. The analysis of variance was generally constructed correctly. Most candidates had a good understanding of contrasts and partitioning of treatment sum of squares. However, marks were lost in part (b). Candidates appeared not to understand in which of the four treatments hay had been cut and/or removed from site. As a consequence, contrasts (1) and (2) were often incorrectly specified. Not all candidates specified orthogonal contrasts as was asked for.

It was pleasing to see a good number of attempts to part (d), with many candidates giving a reasonable interpretation of the residual plots.

Question 2

Part (a) on the principle of blocking in designed experiments was reasonably well done, although some parts were omitted; for example, the degrees of freedom for a balanced incomplete block design.

In part (b) most candidates used the method of least significance differences to compare the treatment means. However, many slipped up with the calculations and forgot to divide the residual SS by its degrees of freedom to give the residual MS ($\hat{\sigma}^2$), and therefore incorrectly concluded no treatment differences. A few candidates performed an overall F test of the null hypothesis for no treatment differences, and just concluded that there were differences between treatments, without exploring the data further. This was not informative.

The circumstances in which a logarithmic transformation of the raw data might be appropriate were quite well understood by most candidates. Few candidates attempted part (b)(iii), but those who did so provided a correct answer.

Question 3

Answers to this question on a 2^3 factorial experiment were somewhat disappointing. The computational parts of this question were generally done correctly. A few candidates

incorrectly identified the degrees of freedom for the interaction sums of squares, assigning 2 df for each interaction. Some candidates inadvertently omitted the SS/df for Hospitals.

Part (ii) was either omitted or not well done. Most candidates appeared not to know how to display the data graphically for multi-factor experiments. Several candidates plotted the mean data for each factor against Hospital 1 and Hospital 2, whereas hospitals were being treated as blocks in the analysis.

The comments and interpretation were generally poor. In part (iii), many candidates simply concluded that the AB interaction and main effect C were significant, but did not attempt to explore the data further. There were some reasonable comments in part (iv).

Question 4

Although this was a fairly straightforward example of first-order response surface designs, few candidates attempted it. Poor answers to part (ii) reflected the lack of understanding of basic regression theory.

Question 5

This question was on stratified sampling with proportions. There were some good performances, but most candidates still got less than half marks.

Part (i) was reasonably well done, although weaker candidates could not explain the circumstances in which stratified random sampling would be better than simple random sampling. Answers to part (ii) were often incomplete. Some candidates could not remember the formulae for the variance of the stratified sampling estimator of a proportion and/or the choice of n for stratification with optimal allocation. Other candidates wrote down the formulae but appeared unable to apply these to the data. Part (c) required candidates to calculate the variance of a sample proportion based on simple random sampling, but few candidates attempted this.

In this survey the gain in precision from stratified sampling was relatively modest. Those candidates who attempted part (d) failed to comment on this.

Question 6

Answers to part (i) were often disappointing. Candidates wrote about stratification and clustering in general, but not necessarily on how and why it might be useful for this survey. In general candidates were poor in considering the practical problems that might arise in developing countries; e.g. in rural areas it is unlikely that maps or sampling frames will be available.

Part (ii) was answered well. Some candidates computed point estimates for the mean number of cattle, which was not asked for; others struggled with the variance calculations, even for simple random sampling. Few candidates commented on the general results for regression and ratio estimators in part (c).

Question 7

Most candidates were able to define a cluster sample in the context of this question. Part (ii)(a) on simple random sampling was generally done correctly but some failed to calculate the standard error of the sample mean. Part (ii)(b) required an understanding of cluster sampling for proportions. Candidates failed to realise this, and used standard formulae for simple random sampling. A common mistake was in the construction of 95% confidence intervals. Often, the upper 5% point of the Normal distribution was used, rather than the *t* distribution, which is more appropriate for small sample sizes.

Answers to part (iii) were satisfactory, though some candidates mixed up the different sampling methods. Stratified sampling requires a complete listing of all patients with trauma events in the past 12 months, which may not be readily available. Many candidates failed to mention this. In general, candidates were poor in considering the practicalities of sampling in institutions such as hospitals.

As improvements, several candidates mentioned sampling in proportional to hospital size or use of stratification by small and large hospitals, which was pleasing.

Question 8

This question on standardised death rates was generally well done. There were a few errors in the calculation of the standardised mortality ratio.

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.