ROYAL STATISTICAL SOCIETY EXAMINATIONS, 2011

REPORTS OF EXAMINERS

GENERAL COMMENTS

This report incorporates the comments made by examiners after marking the papers set in 2011 at all levels of the Society's examinations (Ordinary Certificate, Higher Certificate and Graduate Diploma). We would encourage all candidates intending to take the examinations in 2012 or subsequently to refer to the particular comments on the papers they expect to sit, as this is the primary means by which their examiners can communicate with them. We would also remind candidates that past papers (or specimen papers for new examinations) and reading lists are provided on the RSS website, and strongly suggest that all candidates will wish to make use of these vital resources as part of their preparation.

In 2011, the Ordinary Certificate followed the Higher Certificate and Graduate Diploma in becoming fully modular. Candidates, and those advising them, should be aware of the benefits of a modular structure. Candidates do not need to sit all the modules at a particular level in the same year; indeed, we anticipate that only a small minority of candidates will do so. Candidates are most likely to be ultimately successful in passing at a particular level if they are realistic about the amount of time they have available for study and enter for an appropriate number of modules.

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 2011 echo those made in other 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 2010 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. This year again, candidates lost many marks as a result of their poor skills in algebra and calculus.

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 number of marks allocated to each part of a question, as printed on the examination paper. It is a waste of your time writing a detailed two-page description of some topic, if this can only be awarded two marks.

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

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 strongly advised to show sufficient working to make it quite clear which method is being used.

Be aware of the *RSS statistical tables* that are provided for candidates during all the examinations. This year, several examiners commented on the amount of unnecessary work that candidates had given themselves because they did not simply look up values in these tables. You may freely download a copy of the tables to use during your preparation for sitting the examinations, and you are strongly advised to get to know what is in these tables and how to use them efficiently.

In calculations with several steps, it is important not to round intermediate answers to too few significant figures. For example, if the final answer is to be quoted to three significant figures, then at least four significant figures will need to be retained for intermediate answers.

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 provided 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.

Year after year, examiners comment that many candidates seem a lot more comfortable with calculations and graphs than with discussion or reports.

Applied statisticians need to develop excellent communication skills, so the RSS examinations assess these as well as arithmetic and mathematical skills. You should practise answering discussion questions, possibly using past papers and solutions as a guide; you will find it helpful to talk through your answers with a knowledgeable person, a tutor or a statistician you work with.

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

1. Begin each answer on a new page. (You do **not** need to begin each **section** of an answer on a new page.)
2. Write the number of each question at the top of each page.
3. Graph paper should be attached opposite the answer to which it relates.
4. Enter in the space below (**not** in the side panel) the numbers of the questions attempted. (The question numbers should be written **in the order in which you answered the questions**. Note that the side panel is for the examiners' use only.)

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

ORDINARY CERTIFICATE IN STATISTICS

Module 1 (Collection and Compilation of Data)

General

The motto for this report is, "Answer the question!" In general, individuals and groups of individuals who answered the questions did well. Unsurprisingly, those who did not answer the questions did not score well. It was noticeable that candidates sitting the examination at an overseas centre, as opposed to a UK centre, were more likely to fail to answer the questions set, particularly Questions 4, 5 and 6.

Question 1

Most candidates showed knowledge of the distinction between prices and weights, although not all made the distinction clear in the context of collecting data. Many candidates failed to obtain the final six marks, writing about the qualities of the index or what it might represent instead of its uses.

Question 2

Paradoxically, some rote learners were among the few who earned the two marks for part (ii), which asked for an explanation of sampling bias. Most candidates referred to the possibility that the sample mean might not be the same as the population mean. That this is a truism did not occur. The question expected a reference to the expected value of the sample mean, but the mark was awarded to several candidates who quoted 'systematic', even if that might not have given convincing evidence of understanding.

Very few candidates scored marks on sampling accuracy in part (iv). Most thought that it could be measured by computing the difference between the sample mean and the population mean, even though the whole point of sampling is that the latter is not available.

Part (vi) was answered quite well.

Question 3

Many very poor attempts at the computations were seen. Even more often, interpretational comments were poor in part (v). Comments on the computational results were often not seen, even though the question was asking for an interpretation of them.

Question 4

This question included very specific instructions on what was required, and those following those instructions, and thus answering the question, did well.

Question 5

This was very badly answered by a large number of candidates. The question specifically stated that comments regarding efficiency of service were not required. It specifically asked about data generated by the devices. Yet candidate after candidate made no mention of data; those candidates just did not answer the question.

Question 6

The guidance in this question came within two separate sentences, asking for how to set up the studies, and to compare and contrast them. Far too many students failed to address both sentences, often ignoring the first; they did not answer the question.

Question 7

Some good solutions were seen to this question, although very few candidates indeed made much sense of part (iv). The phrase ‘two separate days’ caused confusion, with students not realising that this would be needed to measure rate of wear. The consequences of different rates of wear, i.e. length-biased sampling, were not appreciated.

Module 2 (Analysis and Presentation of Data)

General

Most candidates attempted all the questions. The best scripts were impressive, showing technical accuracy, good presentation and perceptive comments. The worst scripts showed inadequate levels of preparation, with scrappy solutions, poor diagrams and little or nothing in the way of comments. It was quite common for candidates – even good ones – not to read questions carefully enough; in some cases what they wrote was correct, but because it did not address the precise question asked it could not be given credit.

Question 1

There were very few good explanations of the distinction between continuous and discrete data. Comments like ‘continuous data can take infinite values whereas discrete data can only take finite values’ were common. It was also common for discrete data to be confused with categorical data. The question asked for the answer to be illustrated using the data given. This was quite simple: the original heights would have been continuous data but, when rounded to the nearest centimetre, they become discrete. Most candidates either did not see this or chose to ignore the request altogether.

The stem and leaf diagram was most done well, though it was quite common to omit the key, or to identify the stem as being ‘hundreds’ and the leaves as ‘tens’.

The median and quartiles were mostly found correctly.

Question 2

The arithmetic of large numbers defeated many candidates. A common error was to divide two figures each expressed in millions and to give the units of the answer in millions. The correct calculation gives a figure of about $90,000 per person per year. If this figure is not obtained correctly it is impossible to pass relevant comments on it in the second part of the question.

Question 3

The probability question divided the candidates into three groups. Some made no attempt at all, perhaps knowing that probability is not their strength; some floundered wildly, showing no understanding of the ‘laws’ of probability (not even the ‘law’ that a probability cannot exceed 1); others lapped the question up as an easy source of marks.

Question 4

The arithmetic to complete the table was done well by almost every candidate. Most then got the correct figures for levels of job satisfaction. And the comments were usually broadly correct – though a few candidates, perhaps not wishing to believe the evidence in the data, said that greater academic success *does* bring greater job satisfaction.

Question 5

Everyone knew what to do to produce the pie charts in part (i), but it was evident that many candidates had not brought compasses and protractors with them. Poorly executed and inaccurate pie charts did lose marks.

In part (ii), the inverse ratios to those required were sometimes calculated. Most candidates were able to give good descriptions of the effects of the tax and benefits system.

In part (iii), the calculations were broadly correct, but it was common to give, for example, 7.59 billion instead of 759 billion. This is a further indication that some candidates do not handle large numbers well.

The final part was less well done. The distribution of incomes for the richest 20% of household will have a sharp left-hand cut-off and a long right-hand tail. The positive skew implies that the median will be less than the mean.

Question 6

The arithmetic to find *a*, *b* and *c* was done with mixed success.

The graph in part (ii) was generally well done, though those who were too casual in plotting the data points missed the suggestion that the effectiveness of training may tail off beyond 10 days.

Quite a few candidates made no attempt at finding the regression line. Some drew in a line ‘by eye’, but this received no credit.

The final part was often not tackled at all; candidates seemed to have run out steam.

Question 7

The arithmetic in this question was done well. The comments were less good. In part (i), comments were required on wage increases and staffing numbers in the different grades. Many candidates appeared not to notice the question about staffing numbers. In part (ii), it was not enough to say that the Paasche index takes account of current quantities; it was necessary to say that the quantities, in this case, are staffing numbers.

Question 8

Part (i) was surprisingly poorly done. It was common to say that the data must be monthly because a 12-point moving average has been calculated; this rather misses the point – as does the argument that respiratory disease would have a seasonal variation. For full marks, it was required to say something about the periodicity observable in the peaks and troughs of the data.

It was quite rare to see a *centred* moving average calculated in part (ii).

Many candidates seemed not to know what to do in part (iii). Of those who did, quite a few got the signs of the answers the wrong way round.

In the final part, many candidates wrote about seasonal variation rather than trend.

HIGHER CERTIFICATE IN STATISTICS

Module 1 (Data collection and interpretation)

General

All questions were equally popular. Some candidates lost marks because their written answers were not clearly expressed.

Question 1

The first part of this question asked candidates to form a frequency distribution of 45 observations of a discrete variable, to estimate the mean and the variance of the estimator, and to estimate a proportion and its variance. The second part of the question asked if two samples were simple random. Candidates could form the frequency distribution and estimate the mean. Most could estimate the variance of the distribution, but a sizable number did not continue to estimate the variance of the mean. Estimation of the proportion was fine, but maybe half the candidates did not estimate the variance of the proportion. In the second part, many candidates recognised that the first sample (of two person households taken from a simple random sample of households) was also simple random, but only a minority gave a valid argument that the second sample (of households containing one woman over 30 taken from a simple random sample of households) was not itself simple random. Quite a few candidates tried to relate proportions in the first part of the question to numbers given in the second part.

Question 2

This question was concerned with comment cards left on tables in a restaurant for customers to fill in. It was done fairly well. Candidates could suggest advantages and disadvantages of the method, and potential problems in offering customers the chance to enter a prize draw. Drafts of comment cards to obtain specified information were mainly good.

Question 3

The first part of this question asked about some basic facts to do with social surveys. Candidates appeared to know what is meant by a target population, but many seemed to think that a study population is part of a target population rather than the population studied, for example dwellings might be the study population with a target population of households. A sampling frame could consist of a map rather than a list, but some candidates had some strange ideas as to what a sampling frame is. The problems of coverage when drawing up a sampling frame for a study population were discussed moderately well, but there were few perfect answers. The second part of the question asked for a report on a table containing averages and percentiles for both men and women in two different years. Statistical terms were to be explained, something that few did. From comments it appeared that many candidates did not understand 10^{th} and 90^{th} percentiles.

Question 4

This question gave a grouped frequency distribution of observations on a discrete random variable (time to the nearest minute) where the upper boundary of each class was the lower boundary of the next. Candidates were asked to say why the boundaries were unsatisfactory and to suggest a better set of points and to say what values they would contain. There were some sensible suggestions, but few candidates stated what values their intervals would contain. Candidates were then asked to represent their rewritten distribution by a frequency polygon, and to obtain estimates of the median, the first and third quartiles, and the inter-quartile range. Only about half of the candidates drew a correct frequency polygon. The others drew histograms or cumulative frequency diagrams, and not always correctly. Candidates could calculate the median, but not all attempted to find the quartiles and of those who did many had the quartiles in the wrong class.

Module 2 (Probability models)

General

This year's results compare favourably with those of recent years: overall, the paper seems to have been found rather easier than usual. The increased entry numbers and better pattern of results are gratifying, although there is still a 'tail' of failures including some of derisory quality. Question 1 (on conditional probability and independence applied to intersections and unions of events) and Question 4 (probability calculations involving geometric distributions) appeared to expose common weaknesses among the candidates. In contrast, Question 2 (probabilistic analysis of the distribution of the number of runs in a short binary sequence) and Question 3 (calculations for normal distribution, with brief mention of binomial and Poisson) were generally well done.

Question 1

This question was popular but was rather poorly answered, with an average score below 8/20. Part (i) was generally good, but surprisingly many candidates calculated P(no defects) as $1-\mathrm{P(B)}-\mathrm{P(C)}-\mathrm{P(D)}$. This implicitly assumes that independent defects are disjoint which is, of course, untrue. This error also lost many candidates more marks in later parts.

In part (ii)(a), the probability of 'either or both' of B and C, i.e. $\mathrm{P(B \cup C)}$, was often calculated as $\mathrm{P(B)}+\mathrm{P(C)}-\mathrm{P(B \cap C)}$. Rather few candidates used $\mathrm{P(\overline{B} \cap \overline{C})} = 1\text{-}\,\mathrm{P(B \cup C)}$ along with the independence of D, to solve part (ii)(b). There were very few completely correct solutions to the final part (ii)(c); many candidates failed to attempt anything recognisable as a conditional probability calculation.

Question 2

This question was both popular and generally well answered, with scores averaging over 15/20. Nearly all attempts correctly identified the 10 possible sequences of three + signs and two – signs for 5 marks in part (i), and most attributed the correct number of runs (X) to each sequence. Part (ii)(a) asked for the distribution of X, but a sizeable minority of candidates lost marks by failing to specify this, preferring instead merely to allocate probability 0.1 to each sequence for the purpose of calculating E(X) and Var(X) in (ii)(b). In this final part, a few candidates confused sample and population formulae and divided otherwise correct summations by 10-1 = 9, instead of 10 itself, when trying to find Var(X).

Question 3

All but one of the candidates tried this question on the normal distribution, achieving a creditable average of 13.5/20. Parts (i)(a) and (b) were well done, but the conditional probability in (i)(c) exposed confusion in some minds, shown in errors such as $\dfrac{\mathrm{P(X>189.6).P(X>185)}}{\mathrm{P(X>185)}}$ and $\mathrm{P(185<X<189.6)} = \cdots$, which received zero credit.

Whilst there were many good answers, there was also confusion in parts (ii)(a), (b) and (c), with some candidates submitting for part (a) what would have been correct answers to (c). In part (b), several candidates omitted the combinatorial factor $\binom{4}{2}$, and a few more, perhaps misunderstanding 'exactly two...', included the factor 0.383^2 but not the factor $(1 - 0.383)^2$.

There were several errors in finding the standard deviation of the average height of a sample of 4, presumably reflecting misremembered versions of the $\sigma/\sqrt{n}$ formula. Some candidates also obtained the wrong variance by working with the total $X_1+X_2+X_3+X_4$ and misrepresenting this as $4X$.

For many candidates, the final part (iii) realised a clear 6 marks for Binomial (100, 0.025), Poisson(2.5) and the correct probability (surprisingly, invariably calculated rather than simply found from the *RSS statistical tables* provided). However, several candidates gave the parameters 100 and 0.025 but did not state that Y was binomial, while others stated E(Y) and Var(Y). Most correctly identified the Poisson(2.5) approximation, but several of these represented P(Y > 3) as
1-P(Y = 0)-P(Y = 1)-P(Y = 2). Several candidates used a normal approximation for the probability, which attracted zero credit.

Question 4

Barely 20% of candidates tried this mathematically more demanding question on the geometric distribution: the average mark of about 10/20 largely reflects some very good scores for those who could accurately sum the various geometric series involved, along with some very low marks for those who could not. Throughout the weaker answers, muddle regarding the variable of summation as opposed to the limits of summation, and confusion of '$X > x$' with '$X \geq x$', inevitably lost marks.

In part (iii), the conditional probability yet again exposed weaknesses, and some of those who correctly obtained $\mathrm{P}(X \text{ is odd} \mid X \leq 3) = \frac{p(q+q^3)}{1-q^4}$ had insufficient algebra skills to simplify the RHS to the required answer.

The final part (iv) required summation of a geometric series with 'r' = p(1-p) followed by differentiation and factorisation; this was done well by the mathematically adept minority who tried it, but it was generally not even attempted by other candidates.

Module 3 (Basic statistical methods)

General

Most candidates attempted three questions, though several attempted all four. There were good answers given to all four questions but Question 4 was slightly less popular than the others.

Question by question comments are given below but first here are three areas where many candidates struggled.

(a) In various places in the paper candidates were asked to specify the hypotheses to be tested. A common error was to specify the hypotheses in terms of the data (e.g. writing them in terms of sample means). Remember that a hypothesis is a statement about the population of interest, so sample quantities must not be involved.

(b) Another common problem encountered in hypothesis testing questions was in the conclusion. One does not accept a null hypothesis, neither does one reject an alternative hypothesis. Also, it is important to express the conclusion in terms of the context of the study of interest. So, for example in Question 1, where the variable of interest was plant height, the conclusion should have been given in terms of plant height. Further, if the statistical inference is about, for example, the mean, then the conclusion should also be about the mean.

(c) Each candidate has access to a copy of the *RSS statistical tables* in the examination. These tables are freely available on the RSS examinations web pages and so it is sensible for candidates to familiarise themselves with these tables in advance. Many candidates made poor use of the tables in the examination.

Question 1

(i) This part caused difficulties. All the methods used later in the question depend on the plant heights being normally distributed with unknown mean and variance. Surprisingly few candidates seemed to be aware of this. Many seemed unsure of the difference between an estimate and a parameter.

(ii) This was done well by many candidates, though many wrongly used z-values instead of t-values and some confused the standard deviation with the standard error of the mean.

(iii) There were some good answers but a common error was not spotting that a one-tailed test was required ("... greater mean height ...").

(iv) Again there were some good answers here.

Question 2

(i) As for Question 1(i), many candidates seemed unaware of the normal distribution!
(ii) This part was often well done, but see points (a) and (b) above. Some candidates had difficulty calculating the pooled variance estimate.
(iii) Some candidates had difficulty dealing with the two-tailed nature of the test. Just putting the larger variance in the numerator of the variance ratio and then using the upper 2.5% point of the relevant F-distribution circumvents the need to worry about the lower tail of the F-distribution.

Question 3

(i) This was well done in the main, though some candidates wantonly threw away a mark by calculating the mean to fewer than 4 decimal places even though the question specifically asked for the answer to be given to 4 decimal places.
(ii) This too was largely well done but, having been asked to give the mean to 4 decimal places in (i), some candidates needlessly rounded the mean sometimes to not even 1 decimal place accuracy to do the calculations, thereby generating large numerical errors.
(iii) There were some excellent answers but common errors were not grouping categories with small expectations and not taking account of the estimation of the mean when obtaining the degrees of freedom. It is not appropriate routinely to round expected frequencies to the nearest whole number as this just generates numerical errors.
(iv) This was largely well done.

Question 4

(i) There were very few good answers here. Few candidates could articulate why the Wilcoxon signed rank test was appropriate and fewer still could specify the hypotheses satisfactorily in terms of the differences. Many did not spot that a one-tailed test was required (see lines 3 and 4 of the first paragraph of the question).
(ii) There were several good answers with many candidates performing the ranking calculations well and calculating the test statistic. However, several then struggled to use Table 11 correctly.
(iii) There were many good answers here but common errors were doing a paired t-test rather than a sign test and calculating the wrong p-value (the latter error usually stemmed from not including all the relevant probabilities or from misunderstanding Table 1).

(iv) There were few totally convincing answers. It is better to write a brief paragraph containing the main points than to write down everything you know about the tests in random order.

Module 4 (Linear models)

General

In total, 53 scripts were marked, yielding a satisfactory overall average of 61.5% with 8 failures and 14 distinctions. All four questions had overall average scores of 60% (12/20) or more, suggesting a reasonable level of basic competence in most areas of the syllabus.

However, there were some general weaknesses which should be capable of systematic improvement.

- Several candidates plotted scatter diagrams using dots for points rather than crosses. Dots do not always show up well, so there is a risk of losing marks for apparently plotting too few points.
- Questions 1 and 2 asked for statements of models and assumptions (for ANOVA and simple linear regression respectively), and these were often inaccurately written, with (for example) sloppy, haphazard and incomplete use of suffixes in model equations.
- Somewhat similarly, when candidates sought to state null and alternative hypotheses for significance tests they often stated hypotheses in terms of data values rather than model parameters.
- In Questions 1 and 4 particularly, there was also widespread confusion about the significance level and whether the alternative hypothesis was one- or two-sided.

Question 1

This question on ANOVA was attempted by 47 candidates, who achieved an average mark of 12/20, and of whom about ⅓ answered at distinction standard. However, part (i) (statement of model and assumptions) was surprisingly often very scrappily answered, and indeed was sometimes confused with the simple linear regression model. It was not entirely reassuring that grossly inadequate statements of the model, sometimes an attempt at the data-based partition of the sums of squares not involving the parameters at all, were in several instances followed by correctly calculated balanced ANOVAs in part (ii). Many candidates failed to comment in any way on the clearly anomalous zero observation for method A. Several candidates used $F = 5.10$ in this part, which corresponds to the 2½% level for the (one-tailed) F-test for ANOVA, although the 5% level was asked for. Part (iii) required knowledge of unbalanced ANOVA to take account of the missing observation for method A (previously rendered as 0 in part (ii)); this was less well done, with many candidates making incorrect or incomplete adjustments to their earlier analyses. Several of those who correctly found that removal of the zero destroyed the significance of the between-groups ANOVA failed to note this conflict in the conclusions.

Question 2

This was intended as a straightforward question on simple linear regression. All but 3 candidates tried it, obtaining a good average mark of nearly 13/20; there were 19 answers at distinction level. As with Question 1, the statement of model and assumptions was often inaccurate and/or incomplete; however, the graph and the calculation of the fitted regression were generally good, failures usually being attributable to badly remembered formulae. Part (iii) was weaker: there was a good deal of confusion in applying the F-test for regression (or the t-test for slope which is mathematically equivalent). The final part (iv) was also weak: neither the 'obvious' comment that the model fits the trend of the data with roughly constant scatter, nor the suggestion of both the scatter diagram and the context that the constant term might be redundant, were often made.

Question 3

Only 20 candidates attempted this question, on which the overall average was 12/20. In part (i), many answers noted the low level of explanation achieved by the regression and the global F-test was usually correct; however, the conclusion was sometimes garbled or wrong, e.g. 'good fit' or 'at least two means different' (confusion with ANOVA), rather than that (despite the low R^2) the regression as a whole achieved significant explanation. Part (ii) was of varying quality and sometimes omitted: several candidates appeared not to know how to calculate partial t-values or interpret them. Several candidates omitted to consider the test of the constant term, although this was specifically asked for. Here and elsewhere, some candidates represented significance level as '100% – type 1 error rate', e.g. as 95% rather than 5%. The next part (iii) required consideration of several criteria (rather than just one, such as explanation level, as was typically offered by the more simplistic answers). Good candidates noted primarily that Model 2 had all its parameters significant whereas Model 1 had arguable redundancy between temperature and the constant term (both with partial t not significant). Other points were that: the explanation level of Model 1 is only marginally higher than that of Model 2; the residual mean squares of the two models are very similar; global F for model 2 actually exceeds global F for model 1; and the scatter diagrams of y versus fitted y are very similar. The final part (iv) could be seen as reinforcing these comparisons, as the correctly calculated fitted values $\hat{y}(4.5, 70, 68)$ are also very similar, agreeing to within 2%.

Question 4

41 candidates tried this question on product-moment and rank correlation, achieving an average score of over 12/20 and 15 distinction-level answers. The $s - v$ (or $v - s$) graph was generally accurate (although occasionally with ill-chosen scales) and most candidates noted the increasing trend; however, the pronounced nonlinearity was often ignored, along with the tendency for the scatter to increase with v (or s), and the resulting implication that the product-moment correlation coefficient (pmcc) was therefore less suitable to assess this relationship was seldom if ever mentioned. Both the product-moment and rank correlation coefficients were usually correctly calculated, most errors being due to wrong recall of the formulae. In carrying out the tests of significance, however, many candidates failed to restrict the alternative hypothesis to positive correlation and also failed to realise that the respective tabulated critical values of 0.6319 and 0.6485 at the 2½% level assume a one-sided alternative. Some candidates used the t-test for zero pmcc (which is acceptable but surely more work than looking up the *RSS statistical tables*). Some also used the normal distribution test for the rank correlation (again, more work than using the tables, and only eligible for partial credit as it is an approximate test). Most but not all candidates provided suitable contextualised conclusions. Finally, the rank correlation was usually correctly preferred, as it is valid for any monotonic relationship.

Module 5 (Further probability and inference)

General

This paper includes developments of the probability results and methods introduced in earlier modules, including bivariate distributions and generating functions, and also some of the concepts and methods of statistical inference. In order to be successful in this paper, as well as gaining familiarity with the new statistical and probabilistic ideas, candidates need good skills in algebra (e.g. in evaluating products when constructing likelihood functions) and differential and integral calculus. The standard of attempts at this year's paper was very varied, though 22% of overseas candidates achieved a distinction grade and it was good to see that there were fewer very poorly prepared candidates than last year.

Question 1

A very wide range of answers. The triangular region for the joint density is mentioned in the syllabus, but many candidates could not sketch it. Many candidates made a reasonable attempt at parts (ii) and (iii), though care was needed with the limits and order of integration. Parts (iv) and (v) were generally not so well understood; the distinction in the methods of evaluating $P(Y \le \frac{1}{2} \mid X = \frac{1}{2})$ and $P(Y \le \frac{1}{2} \mid X \le \frac{1}{2})$ needs to be emphasised in class.

Question 2

Not such a popular question, but again very varied answers. Part (a) just tested basic knowledge about the bivariate Normal but was poorly answered in general. Parts (b)(i) and (ii) were generally answered better, though candidates should note that, when they are asked to state a result, there is no need to justify it. A common error in part (b)(iii) was to assert that S and U were independent, which they are not.

Question 3

A more consistently answered question. A surprisingly large number of candidates either gave no answer or a very brief answer to part (a). Parts (b)(i) and (ii) were generally answered well, though in part (i) some candidates forgot to establish that they had found a maximum and not just a turning point. Few knew how to answer part (b)(iii); perhaps something to practice in class!

Question 4

Parts (i), (ii) and (iii) were quite standard and candidates proficient in algebra and calculus got good marks. Parts (iv) and (v) were slightly more unusual, but a few candidates gave good answers.

Module 6 (Further applications of statistics)

General

There were some very good scripts, also a few others where the candidate seemed totally unprepared.

Question 1

Part (a) could be answered as a Randomised Block design or by using a paired t-test. Blocks are pairs, and both methods remove genetic, and any other systematic, differences among the animals taking part.

In part (b), few people really knew what a Latin Square design is and does; those who did know mostly produced good answers although some of the examples were poor.

This year there were several examples of wrong linear models being written down, showing a lack of understanding of the difference between regression and experimental design problems.

Question 2

This was the most popular question, with marks to be gained by doing a standard randomised complete block analysis; this section was well done. The list of assumptions was not always complete. Several candidates who did the analysis were not then able to construct the necessary confidence intervals.

Question 3

This question was the least popular, but there were 4 really good answers among those who chose it. Perhaps charts for proportions are a bit harder to handle than those for means, but both are in the syllabus.

Question 4

In part (a) there were some strange ideas about what exactly regression coefficients tell us; the information about the range of the available data was ignored; dummy variables should be known about in these computer-minded days; and few candidates were able to suggest other variables that could help to improve prediction if put in to the model. Part (b) was answered by several candidates, although some of them were not able to justify their answers with F-tests.

Module 7 (Time series and index numbers)

Question 1

The marks indicate that this is the question which candidates struggled with most. Those who attempted this question showed a poor understanding of prediction.

Question 3

Part (ii) ended with the instruction, "State what the weights are." A description in words was anticipated, but some candidates gave a formula instead; they were not penalised for this interpretation.

Question 4

In part (i), some candidates treated the index as being chain-linked even though the question explicitly gives a single base period. In part (ii), many candidates used the price increases in place of price relatives in the index formulae.

Module 8 (Survey sampling and estimation)

General

The overall standard of answers offered was reasonable. All candidates attempted at least 3 questions. Question 1 was the most popular question, attempted by all candidates, and marks ranged from 11-18. Question 2 on stratified random sampling was the least well done question.

Often, candidates lost marks on the more descriptive parts by not appreciating the practical implications of aspects of the situations described in the questions. Too often, answers were much longer than necessary. Answers need to be more targeted.

Candidates are reminded that they are expected to memorise certain formulae as indicated in the syllabus.

Question 1

Answers to the more descriptive parts were sometimes sketchy. Many candidates discussed the consequences of removals in and out, and deaths on the sampling frame, but not on the conduct of the survey itself.

In part (ii), most candidates could construct the required 95% confidence interval though some slipped up on the calculation of the variance (using n rather than n-1 as the denominator). The question asked why the finite continuity fraction was not needed in the calculation. Many candidates simply stated that the sampling fraction was less than 5%. Only a few candidates justified their statement by explaining that the sampling fracture would be close to unity, and therefore, the finite continuity fracture would have no direct effect on the variance.

In the final part, there was some confusion over the interpretation of 'within 3% of the true value'. The sample size was calculated correctly generally.

Question 2

Answers to Question 2 on stratified random sampling were a little disappointing. Marks were generally below average. The descriptive parts of the question required candidates to look at the survey results, and make an assessment of the merits of using stratified random sampling and optimal allocation. Many candidates did not appreciate this, and offered standard text book responses.

A lot of mistakes were made in part (ii). In addition to mistakes in the calculation of the variance, some candidates estimated the mean rather than the total, while others mixed up the units for number of households, i.e., overlooking 'thousands'.

The final part required candidates to calculate the total sample size and allocation needed to estimate the population total to within d units with 95% confidence. A few attempts were made, but none of those correct. In spite of this, marks were awarded if the candidate had attempted to calculate the sample allocation using an assumed sample size.

Question 3

Candidates were able to explain why the sample selected was a cluster sample, and clearly define the sampling units and sampling plan at each stage. A few candidates thought regions formed clusters, and suggested a 3-stage cluster sample.

Part (ii) required candidates to discuss the practical aspects of setting up such a survey, e.g., sampling a whole class vs sampling individual pupils within a class. Many candidates did not appreciate this, and discussed non-response, which was covered in part (iii).

There were some good responses on part (iii) though many candidates omitted the question on how to use questions about a respondent's background and attitude to investigate the bias.

Question 4

This was the least popular question, attempted by only half of all candidates. However, candidates who answered all parts gained some of the highest marks on the paper. There were good responses to all parts.

In part (i), a common mistake with constructing the 95% confidence intervals was to use the critical point of the Normal distribution, rather than the t-distribution. Given the small sample size for this survey, the t-distribution was considered to be more appropriate.

Part (ii) asked candidates to explain how they would estimate mean weekly expenditure on food (per family and per 2-member family). There were some very good attempts, all mostly correct. Candidates are reminded that the ratio of 2 random variables would be biased in small sample sizes. All candidates who attempted the final part could explain the difference between longitudinal and cross-sectional surveys.

GRADUATE DIPLOMA IN STATISTICS

Module 1 (Probability distributions)

General

Thirty-five scripts were received, of very variable standard; some scored over 90%, others under 20%. Candidates did follow the rubric, and, although it is not specified on the cover sheet, almost always began each question on a new page; this is helpful to markers, but some candidates occasionally continued their answers several pages later, without forward reference.

Question 1

This application of Bayes' theorem was done fairly well. In the standard notation, event A was that the witness stated that a taxi was Blue; the most common (and fatal) error was to take event A as that the witness was correct in making this claim.

Question 2

In this question, some candidates offered little explanation of where their formulae came from: if the final answer is then incorrect, it is difficult to give credit for pages of manipulation. Several weaker candidates treated two different order statistics as independent when seeking their joint distribution, but many candidates gave good answers. A common error, also seen in Question 5, was to fail to specify the range over which an alleged density function applies.

Question 3

The first part of Question 3 was done well by many candidates, with many showing slick manipulative skills, and those who carefully explained what they were doing usually succeeded with the last part. But calculations of the final variance often had errors.

Question 4

This example especially required very careful reading of the question: some candidates assumed that committees were chosen at random from all 26 councillors, but the main substantial error was to seek to use the binomial distribution, not the hypergeometric, in computing the answers.

Question 5

The instruction to sketch the region over which the density was non-zero was meant to guide the candidates into choosing the correct limits for the integrations needed, but many candidates blithely acted as though the non-zero region were a rectangle, and not the triangle they had correctly drawn. Few had the courage to say that E(X)=0 and E(XY)=0, from considerations of symmetry, and the diagram sought in the last part was not done well.

Question 6

This was the least popular question, but several candidates scored very well.

Question 7

This was surprisingly badly done. The ability to express random variables with chi-squared, t and F distributions in terms of a sequence of independent standard normal variables proved elusive, as did the ability to name the distributions of the quantities specified in part (iii).

Question 8

In the first part of Question 8, the instruction to 'outline briefly' was not well understood; answers like "use a Wilcoxon test" do not fit the bill. It was hoped that candidates would describe ways of checking both equi-distribution of values and independence. For the application part, it was essential that candidates explained first how to use random *digits* to obtain suitable values from a continuous U(0,1) distribution. Deriving the distribution function from the density function of a Cauchy distribution was not done well.

Module 2 (Statistical inference)

General

This paper aims to test a range of statistical principles and methods, and their application in simple situations. The likelihood function appears regularly, so it is important that candidates can quickly and correctly derive it, and its log, for a random sample.

Question 1

Some very good answers but quite a few indifferent ones. Two common mistakes when constructing the likelihood were to say that $\sum \frac{(x_i - \mu)^2}{2\sigma_i^2}$ equals either $\frac{1}{2\sigma_i^2} \sum (x_i - \mu)^2$ or $\frac{\sum (x_i - \mu)^2}{2\sum \sigma_i^2}$.

Question 2

A popular, generally well-answered question. A single likelihood $L(\alpha, \beta)$ should be considered, though in this case candidates got away with separate ones. In part (iii), a common mistake was to say $\text{Var}(\bar{T} - \bar{S}) = \text{Var}(\bar{T}) - \text{Var}(\bar{S})$.

Question 3

A popular question, with about half the candidates answering very well and the rest rather poorly. Part (a) was answered reasonably, but candidates should be wary of giving overlong answers to parts carrying only a few marks. For written answers, about two lines per mark is a good rule of thumb. In part (c)(ii), the whole distribution must be shown independent of λ, so the mgf should be evaluated for $\frac{2W}{\lambda}$, not just the mean and variance.

Question 4

A surprisingly unpopular question, given that it is a standard method on a well known application (in acceptance sampling). There were some good answers, but many candidates did not get very far.

Question 5

Another unpopular question, though there were some very good answers. In part (b), in forming the likelihood, it should be noted that it is y_i (the number of unripe mangoes in the ith pack) that has a Binomial distribution not x_i. Also, in part (i) as well as finding the test statistic, candidates should say what the critical region is.

Question 6

Not a popular question, but answered reasonable well by those who did try it. In part (a), though, a number of candidates gave non-parametric tests that were not based on *T*.

Question 7

Quite a popular and well-answered question.

Question 8

Not answered by many candidates. In this type of 'essay' question it is important for candidates to read it carefully, note the sub-questions and make sure that they address each one. In this case the sub-questions are: (i) what is the rationale of likelihood-based inference? (ii) what are the useful properties of estimators? (iii) what are the useful properties of tests? (iv) when is it similar to the Bayesian approach? (v) when is it similar to the frequentist approach. Along the way, likelihood, Bayesian inference, frequentist inference, etc. all need to be briefly described.

Module 3 (Stochastic processes and time series)

General

The two most popular questions were the time series questions, Questions 5 and 7, on the AR(1) and MA(1) models and on Holt-Winters forecasting respectively. The second of these was particularly well answered by a great majority of the candidates, perhaps because it followed a set routine.

Questions 1 – 4

The best answered question on stochastic processes was Question 1 on the branching process. It was particularly impressive that a number of candidates were able to solve the cubic equation in the final part, spotting that one of the roots was 1.

Far fewer candidates were able to supply good answers to the remaining questions. There were no satisfactory answers to the first part of Question 2, which required a careful definition of the notion of intercommunicating states. In Question 4, a number of candidates showed that they did not really appreciate the essential difference between the simple M/M/1 queuing model and the two-server M/M/2 model that was being examined in the question.

Questions 5 – 8

There were a number of reasonable attempts at the applied time series question, Question 6, that involved commenting on computer output and plots, but candidates were not able to provide comprehensively good answers that would have given them really good marks.

Module 4 (Modelling experimental data)

General

Most marks came from the numerical calculations for Analyses of Variance.

Question 1

The extra replication of the Standard treatment was not well dealt with in any part of this question. In (i), proper specification of the model with ranges of all subscripts was absent, and surprisingly few candidates could obtain the least-squares estimates of parameters. In (ii), the double replication of S was not properly catered for in the analysis, although in principle it is no different from a completely randomised scheme (one-way ANOVA) where treatments do not have to have the same replication. Notice also that the comparisons in (a) are not independent of one another, and so there is not a set of orthogonal contrasts which can be dealt with separately. Also in (b), mention should be made of the small percentages among the data set, which might need transformation before analysis.

Question 2

This design consisted of two 4x4 Latin squares, and very few people knew how to find a layout in which two randomly chosen standard squares were put together with adequate randomisation of rows and columns. Part (ii) was better but (iii) showed how few candidates could report on an analysis.

Question 3

Although part (i) was beyond the knowledge of some candidates, the items in the Analysis of Variance were quite often found correctly, useful graphs were drawn and some relevant comments made in (iv). But there was sometimes little idea of how to help plan the next experiment in the series.

Question 4

There were few answers to this question, although a small number of candidates did know about confounding and principal blocks as required in (i), but not all of them could deal with confounding an extra interaction as required in (ii). Residuals as asked for in (iii) were not forthcoming but in (iv) there were some useful comments on what background knowledge would be useful at the planning stage. Some confusion was evident between confounding and fractional replication (which this question did not need).

Module 5 (Topics in applied statistics)

General

A paper like this requires candidates to be familiar with a number of applied methods; the rationale for the method, the basic theory and how it is applied. Overall, the performance this year was encouraging. Some candidates were very well prepared for some aspects of the syllabus but not for others. The topics for questions are not always predictable, so it is important to study the whole syllabus.

Candidates who scored highly were those who had read the questions carefully and answered the specific question that was being asked. In particular, good answers related the theory and analysis to the specific scenario outlined in the questions. Candidates scored good marks when they were precise in their answers and paid attention to detail.

Question 5

The quality of answers to part (i) varied considerably. Not everyone realised that McNemar's Test was appropriate for (ii); the confidence interval in (c) was where most marks were lost.

Question 6

As usual, candidates who attempted a demography question could do the calculations accurately.

Question 7

The theory behind standard sampling processes is not well known. Some marks were gained in the later parts of the question, although in (c)(i) the reasons why Normal approximations can break down often seemed not to be known.

Question 8

Some of the examples given were poor, and in (ii) the necessary information needed was not always made clear. Few people really answered (iii); there are possible systematic or cyclic sample frames that can exist and there is no theory as a basis for analysis unless one treats the data as a cluster sample.