ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

ROYAL STATISTICAL SOCIETY EXAMINATIONS, 2015

REPORTS OF EXAMINERS

GENERAL COMMENTS

This report incorporates the comments made by examiners after marking the papers set in 2015 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 2016 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.

All levels of the Society's examinations are now 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 also 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 2015 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 2015 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 (albeit 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. This point is especially important in the two Ordinary Certificate modules, where there are disparate marks for each question.

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. Examiners have frequently 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 so 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)

In general, the questions were answered well. However, candidates should ensure that they take time to read the questions fully and to then ensure that they answer the questions asked. Also, candidates should note that the breakdown of marks for each question gives an indication of how much detail is required for each question part.

Question 1
The majority of candidates demonstrated a clear understanding of a census, sample survey and sampling frame. However, many candidates did not fully answer part (iii) and did not sufficiently describe steps that could be undertaken to maximise the adequacy of the sampling frame in the context of their chosen survey.

Question 2
This question was answered well. Most candidates demonstrated a good understanding of random sampling. However, in part (iv) some candidates did not answer the question and described a sampling method other than simple random sampling.

Question 3
There appeared to be some confusion between convenience sampling and cluster sampling.

Question 4
This question was answered well with most candidates demonstrating a good understanding of open and closed questions and an ability to design a short questionnaire.

Question 5
Several candidates demonstrated a limited understanding of the differences between missing data and errors in the data and many candidates did not pay particular attention to coding of errors and missing values in part (iv).

Question 6
Most candidates explained the key difference between an experimental study and an observational study, but many did not give enough explanation in part (ii) to warrant 5 marks. Part (iii) was answered well apart from (iii)(d) where many candidates did not explain the steps necessary in processing the data prior to analysis.

Question 7
This question was answered well by the majority of candidates, who demonstrated a clear understanding of the different data collection methods and their advantages and disadvantages.

Module 2 (Analysis and Presentation of Data)

Many scripts showed a pleasing grasp of the concepts and techniques required in this paper. There were very few candidates who seemed completely unprepared for the paper.

Some common problems remain, however.

- Candidates often fail to answer exactly the question asked, and so lose some or all of the marks available.

- Candidates' presentation is sometimes poor. Illegible handwriting *must* result in the loss of marks, and scattering comments or calculations across a cluttered page makes it possible that good work will be unrewarded.

- Statistical diagrams are often poorly chosen and poorly executed. The purpose of a diagram is to convey information accurately and quickly. A diagram that requires a lot of deciphering by the reader is not fit for purpose.

Question 1
The scatter diagram in part (i) was generally well executed, though several candidates chose a poor scale for the vertical axis. A false origin is pretty much essential here if the information is not to be lost.
The interpretation in parts (ii) and (iii) was often good.

Question 2
The calculations in parts (i) and (ii) were found easy in general.
In part (iii), many candidates simply put the figures in the table into a narrative when the sort of interpretation required was to use the figures to build up a picture of the different attitudes to defence spending in the different regions of the country.

Question 3
As with Question 2, there was a tendency here for the figures in the table to be regurgitated rather than interpreted. This was also a question in which candidates sometimes strayed from what is asked for. So, for example, part (iii) asks how a country's road safety could be assessed 'using these data'. It is not appropriate, therefore, to talk about road safety campaigns that countries might introduce.

Question 4
The first three parts of this question were answered well. It was pleasing in part (ii) to see candidates interpreting the variances as well as the means.
Part (iv), however, was quite often poorly answered. Candidates talked in very general terms about the correlations being positive or negative, weak or strong. What was required was an interpretation of the correlation coefficients in terms of the effects of the treatment on blood pressure.

Question 5
As usual with probability questions, this polarised candidates. Some got full marks on what is a very standard application of the probability laws. Others, clearly ill at ease with probability, struggled to write anything sensible.

Question 6
This was, for the most part, a very straightforward question for the majority who had thoroughly learned about price relatives and indices. A substantial minority, however, showed that they had picked up a smattering of the ideas and terminology, but not enough to carry out any of the calculations correctly.
Part (ii), on annualised rates of inflation, proved to be beyond most candidates. It was rare to see any attempt to take 10^{th} roots, but common to see division by 10.

Question 7
There were some good solutions to parts (i) and (ii) this question, with candidates correctly explaining that seasonal variation having an increasing amplitude is best handled with a multiplicative model. Surprisingly, though, the terminology of time series often became muddled with statements like 'the trend is for the seasonal variation to increase' quite common.

Quite a few candidates didn't understand centred moving averages, and many were unable to use the figures in Table 2 to demonstrate that the multiplicative model works best.

Question 8
In this question, parts (i) and (v) were done well, but candidates performed less well on the rest.
In part (ii) there were very few convincing demonstrations that the given figure of 82% was consistent with the data in the question. More seriously, perhaps, the diagrams in parts (iii) and (iv) were often poor. A simple bar chart (not a histogram!) serves well in part (ii). In part (iii) the best diagram is a set of line graphs on the same axes showing, for each region, the internet usage figures in 2005, 2010 and 2013.

Question 9
This was found to be quite straightforward by most candidates. The outlier (128) was identified and explained in terms of its distance from the rest of the data. (Some candidates spotted that 154 is an outlier; that answer was accepted even when 128 was overlooked.) The explanations for the outlier in part (iii) were suitably imaginative, and in part (iv) most candidates correctly stated the effects on the summary statistics of removing the outlier.

HIGHER CERTIFICATE IN STATISTICS

Module 1 (Data collection and interpretation)

Many candidates displayed a good grasp of relevant issues in collecting and interpreting data. However, some scripts were difficult to read, either because candidates' hand-writing is poor, or because it was not clear which part of the question an answer referred to.
Some weaker candidates ignored the detailed requirements of the question in favour of writing as much as possible of a general nature, hoping thereby to pick up marks. Generally speaking, this is not a successful approach.

Question 1
This was the least popular of the four questions, perhaps because it was the least structured question and so required the candidates to tackle a rather open ended task.
Weaker answers tended to regurgitate the figures provided in the table. Extracting figures from a table and putting them into a narrative form makes the information *harder* to absorb, not easier. A better approach is to start with the major similarities in mode of transport and then to identify the most interesting differences.
Many candidates ignored the instruction to incorporate appropriate diagrams (plural) into their report. The diagrams that were provided were often of poor quality.

Question 2
The calculations required in parts (i) and (ii) of this question were generally done accurately, though some candidates seemed to make very heavy weather of them. In part (iii) the explanations of how a sample of size 1 has a zero standard deviation were generally good.
In part (iv), many candidates appeared to misunderstand the question. The requirement was to look at systematic differences in two distinct categorisations: male versus female, and learner type. Candidates commonly conflated these two, analysing gender differences within learner type. This, of course, led to lower marks.
In tackling part (v), many candidates missed the causal link in the lecturer's claim, focusing instead solely on the numerical data.

Question 3
This question was generally tackled well. In part (i), most candidates were able to make constructive criticisms of the preliminary questionnaire and to suggest improvements. Issues commonly identified included relevance, repetition, unclear wording, negative wording and the lack of any open ended questions.
In part (ii) most candidates made a good job of providing an introduction that emphasised the importance of health and safety and encouraged workers to complete the questionnaire.
The answers to part (iii) mostly identified the problems that could arise with a low completion rate. However, many of the suggested ways of getting a higher completion rate conflicted with the need for confidentiality in a questionnaire in which management might be criticised.

Question 4
This question was the most popular and also the best answered of the four. Most candidates appeared to have learned the topic thoroughly and they wrote confidently about the different sampling methods. It was noticeable, however, that answers were often given at far greater length than was required. In some instances this led to candidates running out of time to complete the paper.
Part (ii), however, was generally well done.

Module 2 (Probability models)

The exam overall was done well. Here are some comments on individual questions.

Question 1
This was done the least well. In (a) part (ii)(C) many of the candidates did not understand the B is a subset of A notation. In (b)(ii), the results of the second test required updating the probability of the disease again given the new information, so reapplying Bayes theorem. This confused some candidates. In the last part of (c) part (ii) they needed to calculate the conditional probability of C drawing the game given that M has won the championship. Some candidates omitted to include all the possible cases.

Question 2
This question was done very well in general. For 2(a), in an explanation of why a continuity correction is used, often was omitted the fact that for a continuous distribution the probability of X equalling a single value is zero, so we need to increase the area under the curve to allow for this as this is not the case for a discrete distribution. The explanation of when approximations should be used did not always include the need for a large sample size. I was surprised that since earlier in the question the condition for needing a continuity correction had been discussed, this was often not utilised in the Normal approximation.

Question 3
This question was extremely well done. In part (ii), the probability Y lies between 0.5 and 1.5 could easily be calculated using F(1.5)-F(0.5), which saves a great deal of work, rather than integrating the pdf.

Question 4
This question was mostly done well. In part (i) several candidates ignored the fact that the means and variances of each X_i are different, so
$E(T)= \mu_1+\mu_2+ \ldots+\mu_n$ etc., and this was needed throughout the question. The next two sections of the question were done well, but the last part required expressing the mean and variance of the weight of the books in terms of n, the number of books on the pallet, and then finding the probability that this weight is greater than 100 is less than 0.01. This means equating the Z values with the appropriate Z value in the tables. Some of the candidates did not have any idea how to tackle this.

Module 3 (Basic statistical methods)

Question 1
This was done by the vast majority of candidates. The standard deviation was incorrectly calculated on a small number of papers. However, the confidence intervals and hypotheses tests were on the whole well done, perhaps slightly less so for the standard deviation than for the mean. As the sample size is small, the *t* distribution should be used for inference about the mean rather than the normal distribution. Candidates should be careful to present clear conclusions at the end of the hypothesis tests – stating whether or not the null hypothesis should be rejected or not is not sufficient. Non-significant test statistics indicate that the null hypothesis should not be rejected, rather than that it should be accepted.

Question 2
Part (i) was well done by most candidates. There seemed to be confusion about looking up chi square values in tables, where some candidates halved their required significance level in order to look up values. This is not necessary as only large values of the test statistic lead to reject of the null hypothesis. The confidence interval for the proportion in (ii) was (0.1425, 0.2317). In the ensuing comments, it was important to state that if the distribution was uniform, one would expect 2/7 (0.2857) of births to be at weekends, whereas the confidence interval does not contain this value, indicating that fewer births than expected are at weekends. The hypothesis test in (iii) was for a difference of two proportions and it is

important to remember that the formula for the standard error of this difference is different for hypothesis tests to confidence intervals. Many candidates lost marks through using the wrong formula for this quantity.

Question 3
Part (i) required the form of a confidence interval for the difference between two means for the case where the variances are equal. The formula for this was not always well known, and a common mistake was to write 1/n as 1/9 rather than 1/10. The nonparametric test in (ii) was well done by most of those who attempted it, although a few tried to pair the data values and then apply the wrong test. Solutions should state specifically that the **lower** of the two rank sums is used as the test statistic. The Wilcoxon test is easier to perform and it does not depend on the assumption of normality, however it is not as powerful as the corresponding *t* test.

Question 4
The null hypothesis in (i) was not always well expressed, it should mention no association between charging discrepancies and whether or not an item is on special offer. However, the calculations were well done, and conclusions were stated clearly. Part (ii) began by asking for the given table to be adapted, although not all candidates actually wrote down the new table. The confidence interval for the difference in two proportions required in (ii) was not generally well done as candidates seemed to have some difficulty in knowing what to do (the writing down of the table initially was intended to make this clearer). The results in (i) and (ii) do actually appear to contradict one another. This is because the difference in price accuracy found in (i) was largely due to the differences in numbers of under/overcharged items whereas in (ii) those two categories were amalgamated.

Module 4 (Linear models)

Overall the exam was done very well. Here are a few remarks on each question.

Question 1
This was done the least well. Parts (i) and (iii) were done well. It was disappointing that many of the candidates were unable to do the derivation of the least squares estimates of the simple linear regression. The analysis of variance was also not done particular well by some, thus leading to a loss of 15 marks on question 1. The simple linear regression question although straightforward, could not be done if candidates have not learnt the derivation or analysis of variance table for that model.

Question 2
This question was done very well in general. In part (i) very few of the candidates knew a suitable constraint on the parameters. The final part required a t-test to test the difference between two of the carpet products. There were various different attempts at what was an appropriate s^2 for the variance of the difference in means. Also in this part of the question the 0.5% significance level was often not used properly or just ignored.

Question 3
In part (i) some candidates did not write the formula down for the pmcc, but some wrote down the formula later so I gave them the marks. They clearly did not understand that “define” means write down the formula. Working out the pmcc for a linear transformation of the x variable was not attempted by many. In the last two parts the pmcc and the Spearman’s rank correlation coefficient were generally done well, but several of them did not use the tables specifically for the pmcc and Spearman’s rank correlation coefficient.

Question 4
This question was mostly done very well. At the start of the question the full error assumptions for the multiple linear regression model were not always stated fully. In part (i) the null hypothesis should have been stated in terms of the regression coefficients. In part (ii) for the confidence interval for the coefficient representing Head, the correct standard error was not always used.

Module 5 (Further probability and inference)

Question 1
This was generally quite well answered, though a surprising number of candidates did not answer the basic questions in the first paragraph fully. Some candidates incorrectly thought that a correlation of -1 meant that X and Y were proportional. Much of the rest of the question revolved around the result that if X and Y are bivariate Normal then any linear function, such as $aX + bY$, has a univariate Normal. Some candidates forgot the covariance term in the result: $\text{var}(aX+bY) = a^2\text{var}(X) + b^2\text{var}(Y) + 2ab\ \text{cov}(X.Y)$.

Question 2
This was a standard question, with some very good answers. Some common occurrences concerning likelihood were (i) not adding a subscript i to y (or k) when writing out the likelihood, and (ii) using the likelihood rather than the log likelihood when finding the mle (not a mistake, but likely to lead to a more difficult solution). Several candidates did not know how to use pgf's to find the distribution of a sum of random variables and some did not mention the one-to-one correspondence between discrete distributions and their pgf's.

Question 3
It was good to see that the differentiation was carried out correctly by most candidates. Once part (i) was answered, part (ii) should have been straightforward, but many candidates went on the wrong track.

Question 4
This question is of a type that has been regularly set before, so it was surprising how poor nearly all the answers were. Many candidates seemed to be very confused about this area of the syllabus.

Module 6 (Further applications of statistics)

In many cases the candidates coped adequately with number-crunching questions, such as the stepwise regression and ANOVA, and the simple algebra for the quality control question. However, very few candidates performed well in the more explanatory questions, such as experimental design or residuals, or comments and critical assessment of the models used in the number-crunching. This has been the pattern for the last 2 years as well. The question paper has a substantial amount of more discursive questions of the latter type so that the overall performance was therefore poor.

Module 7 (Time series and index numbers)

Overall this paper was answered well, with most candidates showing ability and knowledge in both subject areas. Candidates were frequently noticeably better at the calculations than at describing properties or interpreting results.

Question 1
This question covered moving averages and their properties. Most candidates could define the simple exponentially weighted moving average and carried out the required calculations accurately. On the other hand there were surprisingly few accurate definitions of a symmetric moving average even though notation for the weights was given. The second half of the question was on the properties of these two types of moving average and here many of the candidates were on less sure ground.

Question 2
This question covered ARIMA modelling, with particular emphasis on interpreting computer output. Only a small number of candidates chose it, but they were reasonably successful.

Question 3
This question applied index numbers to hospital death rates, a non-economic but topical context which might have proved challenging. In fact Q3 returned the highest average score. The question gave ample advice on how to relate the health statistics such as death rate to the more familiar economic quantities like price. Most candidates successfully interpreted the question and carried out the calculations accurately.

Question 4
This question concerned the definitions and relationships between economic indices and the practice of ‘chain-linking’. Although there were some excellent attempts there were also several low marks. In particular some candidates misunderstood the definition of chain-linking, which was crucial to the numerical example.

Module 8 (Survey sampling and estimation)

The overall standard of this paper was good, with a high proportion of candidates scoring marks in the range 30 to 40. All candidates attempted 3 questions, and all 4 questions attracted a similar number of attempts.

Question 1
Answers to part (i) were rather disappointing. Typical problems such as locating the sampling units from the air, the quality of the aerial search, and the accuracy of counting animals in the photos, etc, were seldom noted. Some candidates wrote about the errors caused by major movements of herd taking place during the survey, which was pleasing.
Most candidates were able to estimate the total number of caribou in the herd and calculate the estimated variance V using the formula given in part (ii). However, there was some confusion with the standard error to be used in the 95% confidence interval, with some candidates using $\sqrt{\frac{V}{n}}$, rather than $V^{½}$. A few candidates lost marks through calculating an unweighted mean of the stratum means, as an estimate of the mean number of caribou.
The merits of using stratified sampling in part (iv), and using optimal allocation in part (v) were well understood, but sometimes candidates wrote too generally, and not in terms of what the data showed for this survey.

Question 2
This question related to a health survey; candidates seem passionate about the subject matter and provided some good discussion, which was pleasing.
There were some good answers to part (i), but sometimes candidates did not give enough pertinent points to obtain full marks (or wrote extensively on one aspect). The advantages of collecting data using a diary were more extensively discussed than the disadvantages. Aspects such as the cost of executing a diary, its limitations in some population groups, and people’s unwillingness to participate in the first place, were seldom mentioned.

The concept of bias in part (ii) was well understood, but not everyone was familiar with the concept of imputing item responses by matching non-response people, for these factors and lifestyles, with those who have replied.
There were good responses to part (iii) on questionnaire design. Most candidates offered sensible alternatives to the proposed question.

Question 3
This question was based on simple random sampling. The data were presented in a frequency table. Many candidates struggled with the calculation of the standard deviation required in part (i)(a), and lost marks. Everyone got part (i)(b) correct.
Only a few candidates stated that the number of persons per bedroom in part (ii) was a ratio estimator. Note that this follows, since the number of persons and number of bedrooms varies by household. Candidates' discussion of the practical issues in defining "households", in defining "overcrowding" and including information from single-person households, was satisfactory.

Question 4
There were good marks to all parts of question 4.
Part (a) was descriptive; there were some good answers, but sometimes candidates did not give enough pertinent points to obtain full marks. Some candidates did not appreciate the sensitivity of the topic and how that would influence the choice of data collection method used.
Part (b) tested candidates' understanding of *cluster sampling*. A few candidates were unable to define one- and two-stage cluster sampling.
Part (c) on determination of sample size was done well, with most candidates scoring full marks. However, the discussion of practical issues in planning this survey was rather disappointing. Issues of defining a student's average weekly earnings, and how to locate the target population when only a list of registered students is available, were often omitted.

GRADUATE DIPLOMA IN STATISTICS

Module 1 (Probability distributions)

Candidates generally performed well in this paper and several of them achieved excellent marks. Attempts at Question 6 were particularly good, but the average marks for Questions 1 to 5 were also all above 10/20.

Question 1
This question examined the Law of Total Probability and Bayes' Theorem. 90% of candidates attempted it and they gained more than half marks on average. A few candidates, surprisingly, failed to write down the Law of Total Probability correctly for part (i). A number of candidates did not realise that part (v) was an example of a Normal approximation to the binomial distribution; the word 'approximation' in the question should have made this clear.

Question 2
Here candidates had to work with a bivariate hypergeometric distribution and its marginal distributions (which were, therefore, also hypergeometric). About $^2/_3$ of candidates answered this question and their median mark was well above 10/20. Some candidates did not put the correct values into the formulae given in part (v) in order to find the marginal expected values and variances of the random variables. In part (vi), a substantial number of candidates did not point out that $X + Y + Z$ must always equal 5 so its expected value must be 5 and its variance must be 0.

Question 3
About 60% of candidates attempted this question, on discrete random variables in the context of a game of chance. They gained more than half marks on average. Some candidates appeared to struggle to find the variances required in parts (ii) and (iv) (c); this was partly because they were using the formula $E\left[(X - E(X))^2\right]$, rather than $E(X^2) - [E(X)]^2$ which is usually more suitable for arithmetic calculations.

Question 4
This consisted of 4 short proofs relating to the Poisson distribution and its relation to other distributions. About 60% of candidates attempted it and gained just over half marks on average. Candidates would have found it helpful to write down systematically the information given in part (a): that X has a Poi(λ) distribution and Y given X has a Bin(x, θ) distribution. Mistakes made in extracting this information cost some candidates a lot of marks.

Question 5
Almost all candidates tackled this question, on beta distributions in one and two dimensions, and many of them made excellent attempts at it. Candidates lost most marks by mis-specifying the limits on the integrals required at the start of part (ii) to find the marginal distributions of the two random variables. To avoid such problems, candidates should always sketch a rough diagram showing the joint range space for examples such as this.

Question 6
This question was chosen by only 30% of candidates, but their attempts were almost uniformly very good. They recognised that this question was based around the Box-Muller method for simulating from the Standard Normal distribution, and were clearly adept at working through the standard proof.

Question 7
This question examined the moment-generating function of the Normal distribution. About $^2/_3$ of candidates attempted it, but with mixed success. Some candidates struggled with the calculus required to derive the moment-generating function in part (i), but this is a standard proof for this paper. A majority of candidates had problems working through part (ii); again, they seemed to struggle more with the calculus than with the underlying concepts.

Question 8
Only a small number of candidates answered this question, on simulation using the inverse c.d.f. and rejection methods. Generally, answers were rather poor and only one person made a serious attempt at part (iii). The rejection method is part of the syllabus for this paper, so candidates should expect questions on it from time to time.

Module 2 (Statistical inference)

Candidates mainly attempted questions 1 to 5, presumably because most revision time had been spent on these topics. However, by neglecting parts of the syllabus they may be reducing their chances of passing.

Question 1
This was a popular question, usually done at least reasonably well.
When forming the log likelihood, it saves trouble later to recognise which terms can be considered "constant" as early as possible.
Always check that the turning point is a maximum when finding a maximum likelihood estimator (mle) – usually by showing that the second derivative of the log likelihood is negative at the mle.

Question 2
Popular and usually well answered apart from part (iv).
Part (ii): in order to show that Y is pivotal, it necessary to note that it is a function of the unknown parameter as well as showing that its distribution does not involve the unknown parameter.
Part (iv): the mention of the word "power" caused some candidates to start trying to use the Neyman-Pearson lemma, despite the form of the test being specified.

Question 3
Popular and usually well answered.
In describing the Newton-Raphson method it is necessary to give some words of explanation as well as a formula; in particular, all terms must be defined.
Part (iii): under regularity conditions, the mle is asymptotically unbiased, so there is no need to try evaluating the expectation of the mle.

Question 4
Popular, but with variable answers.
Part (ii): candidates should not spend too much time on this type of problem if after a couple of minutes they are not making progress.
Part (v): a standard question but with relatively few good answers.

Question 5
Popular, but with relatively few good answers.
Part (iii): it is important to explain clearly how $\log \lambda$ is to be found; this will attract marks even if the precise form given in the question is not obtained.

Question 6
Not a popular question although it is quite standard.
Common errors in describing the Wilcoxon test were to say that ranking is based on Xi-m0, rather than the absolute value of this, and not to say that the test statistic is equal to the minimum of the sum of negative ranks and positive ranks (or something equivalent).
A number of candidates seemed to be unaware of how to use this test to find a confidence interval.

Question 7
Moderately popular, but with few good answers.
Several candidates missed steps out in attempting to find the jackknife estimator.
Part (v): many candidates found the Bayes estimator of λ instead of θ.

Question 8 had very few attempts.

Module 3 (Stochastic processes and time series)

This year there were 5 questions on stochastic processes and 3 on time series. As usual, the time series questions were the most popular. Indeed, all candidates attempted question 6 on the infinite moving average representation of an AR(2) process. Most candidates were able to verify stationarity through examination of the roots of the autoregressive characteristic equation, but many candidates were not able to successfully atttempt the rest of the question. Many candidates made a reasonable attempt at the more practical questions 7 and 8, on ARIMA modelling and exponential smoothing, respectively, but it was disappointing how many in question 7 were not able to discuss clearly the use of the acf and pacf to identify appropriate models to fit to the data.

The most popular of the stochastic processes questions was question 2 on the properties of a 4-state Markov chain. Many candidates were able to calculate successfully the 2-step probabilites in part (ii), but it was disappointing how few had a sufficient understanding of the concepts of transient and recurrent states required in part (i). The least popular was the other question on a Markov chain, question 3, but those candidates who attempted it gave some very good answers, including the evaluation of the stationary distribution.

What was a fairly routine question on a branching process, question 1, was attempted by fewer candidates than might have been expected. Question 4, on the properties of a continuous time Markov chain model, was very well done by a few of the candidates, but question 5, which involved concepts that arise in the discussion of the embedded Markov chain for an M/G/1 queue, was, perhaps unsurprisingly, the one that produced the fewest good solutions.

Module 4 (Modelling experimental data)

Good marks were generally achieved where candidates addressed all parts of a question. In design-focussed questions it is generally important to be able to construct the sets of treatments (in blocks) to be considered, and to check that, when constructed, these sets do have the required properties. For questions focussed on the output from analyses, the key element to the question is usually the interpretation of the presented (or calculated) results, and this interpretation needs to go beyond a simple description of what has been presented, ideally providing inferences and conclusions in a non-technical language.

Question 1
This is the kind of design question that should be standard work for the applied (consultant) statistician, requiring the identification of possible designs, and the selection of the design that meets particular criteria. Most answers failed to get particularly high marks because one or more components of the question were not answered. Many answers failed to explain the second relationship about the numbers of pairwise comparisons, and most answers failed to state that all five parameters needed to have integer values to identify a BIBD. The calculations of block sizes and numbers of blocks were generally good, though many answers failed to identify the maximum block size as being one less than the number of treatments (by definition for a BIBD). Some answers failed to identify the non-integer number of blocks as the problem with $\lambda = 2$. Comparisons of the possible designs generally considered a large number of blocks to be a problem, whereas the block size is the most important consideration. Most answers failed to include the construction of the $\lambda = 3$ design, possibly reflecting that the approach to doing so by hand is not generally used, relying too much on computer implementations.

Question 2
This is the type of design challenge that applied (consultant) statisticians are often faced with – to identify the most appropriate form of design to make most effective (efficient) use of the available resources while taking account of various practical constraints. While there were some good answers, many answers were far too focussed on the equations for constructing sums of squares and parameters, rather than on the practical issues associated with identifying the most appropriate arrangement of treatments to cope with the scenario. There was some uncertainty about what should be included in the dummy ANOVA tables, with a number of answers failing to identify the numerical values of the degrees of freedom – this is the most useful information when comparing candidate designs in the absence of any information about the variance components. Some answers failed to identify the three randomisation steps for the randomisation of a Latin square design – there is no need to randomly select from all possible Latin squares of the appropriate size, but to first select from the set of standard squares (with first row and column in alphabetical order), then to randomly permute both rows and columns (separately), and finally randomise the allocation of treatments to codes. Many answers captured the essence of the split-plot design that was appropriate for the final part of the question, though often without suggesting a sensible blocking structure for the main plot treatments, such as a pair of 3-by-3 Latin squares.

Question 3
This is a common type of design problem, requiring the statistician to balance various different constraints in constructing a suitable design, but was not attempted by many candidates. Answers showed a general understanding of confounded and fractional factorial designs, though the distinction between confounding contrasts and fractionating contrasts is clearly not understood – confounding contrasts identify groups of factorial combinations to be included in blocks, while fractionating contrasts identify the aliasing structure. Most answers correctly identified fractionating contrasts for a quarter fraction, though often the justification for particular choices was not clear – choosing two 4-factor terms creates issues with the later choice of confounding contrasts. The identification of the aliasing structure was generally good, though with a few careless errors. The principle of identifying the principal fraction is generally understood, though many answers failed to identify that a quarter fraction would contain 32 factorial combinations. No answers successfully completed the confounding of these 32 factorial combinations into 4 blocks of size 8.

Question 4
On average the best answered question, though there were some very good and complete answers and others that displayed a lack of understanding of these analysis methods. While many answers identified the hat matrix as containing the leverage values on the diagonal, very few answers explained how these values could be calculated more simply for a simple

linear regression model. Most answers correctly identified that leverage was not associated with the response but just with the explanatory variables, with influence associated with the response variables, but some confused leverage with points having large residuals. Sketch plots were generally good, though many would have been clearer with the addition of fitted regression lines – often the "low leverage, high influence" scenario was not particularly clear. Most answers showed a good understanding of how to calculate both the coefficient of determination and the adjusted coefficient of determination, but the explanations of why the former has to increase as the number of explanatory variables increases (residual sum of squares cannot increase) while the latter will only increase if an additional variable explains a sufficiently large proportion of the residual variation. Most answers described how to calculate Mallow's Cp, but very few demonstrated how this statistic could contribute to a simple graphical identification of good models. The interpretation of the fitted model was generally poor, though some identification of possible collinearity issues causing the counter intuitive parameter estimates.

Question 5
Another popular and generally well-answered question, with a few almost complete answers. Common errors were to fail to include the block term in the construction of the ANOVA able, and to fail to subtract the main effect sums of squares when calculating the interaction sums of squares. Relatively few answers commented on the (lack of) differences between blocks, but otherwise interpretation of the ANOVA table was good. Most answers demonstrated a good understanding about the required conditions for orthogonal contrasts, though few answers provided sensible sets of coefficients for linear and quadratic contrasts for five equally spaced levels. Calculations of sums of squares for contrasts were generally good, though often for the wrong set of contrast coefficients, and very few answers provided the values of the contrasts as well as the sums of squares. Interpretation of the tests of the contrasts was generally poor, though possibly as a result of the incorrect identification of contrast coefficients, and few answers provided a sensible assessment of evidence for lack-of-fit for the linear and quadratic models.

Question 6
Another popular question, though many answers were rather incomplete, often failing to provide any interpretation of all of the presented analysis output. A number of answers failed to note the underlying assumption of homogeneity of variance when explaining about variance stabilising transformations. Where answered, the derivation of the relationship to define a variance stabilising transformation was generally good. There was some confusion about the two situations presented in part (iii) – a few answers were the wrong way around! Many answers identified one of the plots (variance v. mean; standard deviation v. mean) as clearly being more linear than the other, whereas these data are rather inconclusive. Most answers also failed to note that residual plots for the square root transformed data still seem to show the variance increasing with the mean. A few answers noted the non-significance of the interaction for the analysis of the transformed data, and hence the simpler interpretation. Relatively few answers identified that a Poisson GLM would be a good alternative approach, allowing an assessment of the quantitative response to concentration and the comparison of regression lines between species.

Question 7
The least popular question, and generally rather poorly answered, though there were a couple of very good and almost complete answers. Some answers were confused about the difference between general linear models and generalized linear models, often failing to include that the response variable is assumed to follow a distribution from the exponential family. None of the answers provided a complete description of the method of iteratively reweighted least squares, most failing to appreciate both the iterative and reweighting aspects. Most answers failed to construct the accumulated analysis of deviance tables, and where appropriate comparisons of models were reported, often these used an (approximate)

F-test rather than comparing the change in deviance between models with the critical values of the appropriate chi-squared distribution. A formal test of the residual deviance, to check for lack-of-fit, was rarely included. There were a few good explanations of the derivation of the parameterisation in terms of the LD50 (by identifying the LD50 as giving a response of 0 on the logit scale), and also a few good explanations of why it was appropriate to describe the results in terms of relative efficiency for the fitted parallel lines.

Question 8
A relatively unpopular question, with a couple of very good and complete answers, but also a number of answers showing a complete misunderstanding about the analysis of contingency table data using a GLM. Most answers failed to identify that the data were multinomial in nature, but that the data can be analysed using a Poisson GLM because the distribution of counts across the categories of one factor is conditional on the marginal counts. Many answers failed to identify that the saturated model in this scenario has a deviance of zero on zero degrees of freedom, so that the residual deviance for the final presented model provides a test of whether this more complex model is needed. Many answers failed to describe the baseline model as representing the overall distribution of counts across the two response (respiratory infection) categories, plus the distribution of the counts across the different combinations of levels of the potential explanatory factors. We are interested in how these explanatory factors affect the distribution of the counts for the response categories, not in the distribution of counts across the different combinations of explanatory factor levels. In considering a stepwise approach it is important to consider the number of degrees of freedom (and hence significance) associated with each change in deviance, not just the magnitude of the change in the deviance (i.e. choosing the model with the smallest residual deviance at each step – though in this case the same “best” model was still identified with this approach). Some answers incorrectly tested the mean deviance associated with each change, rather than the total deviance. There were some good attempts at interpreting the final model, though most failed to revisit the table of counts to determine which levels of particular factors increased the risk of respiratory infection – this is what a lay reader would want to know, not just which factors were significant!

Module 5 (Topics in applied statistics)

Candidates generally performed well in this paper and several achieved excellent marks in it. Many candidates seemed more comfortable carrying out mathematical or arithmetical tasks than engaging with a practical context or expressing their ideas in written form. Candidates should practise writing short notes on important topics from the course and writing brief conclusions from analyses they carry out during their preparation for the exam.

Question 1
This question examined discriminant analysis, in a real-life context. About half of all candidates attempted it, achieving over half marks on average. Several candidates were confused about the comparison of the linear and quadratic discriminants, though most realised that the cross-validation results indicated that the quadratic discriminant had a tendency to overfit to the observed data.

Question 2
In this question, candidates had to write short notes about various aspects of cluster analysis. Only a minority of candidates answered it and few of them did so successfully. In part, this was a particular example of the general difficulty many candidates had in expressing themselves clearly in writing. Over and above that, several candidates gave the impression of knowing only about some aspects of clustering, not all the topics listed in the syllabus.

Question 3
Almost all candidates attempted this question on survivor functions and the Weibull distribution. Very many of them scored excellent marks for it.

Question 4
This was the second question on survival and reliability, dealing with the Cox proportional hazards model in a practical context. About $^{3}/_{4}$ of candidates tried this question and they gained over half marks for it on average. Many candidates did not answer all the parts of the question. In part (i), almost no-one made an effort to interpret the terms of the Cox proportional hazards model in the context of the question. In part (ii), a large proportion of candidates did not explain how to combine an estimate and an estimated standard error in order to arrive at a statistical conclusion. In the same part of the question, very few candidates mentioned the direction of any effect: for example, does the evidence suggest that having the new treatment increases or reduces the hazard of infection?

Question 5
This question presented data from a new diagnostic test. Candidates had to discuss its sensitivity and specificity at various threshold values, then construct and interpret a ROC curve. About $^{2}/_{3}$ of candidates attempted it, with mixed results. Many attempts at part (i) failed to discuss the likely usefulness of this particular test in light of the calculated sensitivity, specificity, PPV and NPV; no marks were given for comments about the properties of diagnostic tests in general. In part (iii), only a small number of candidates attempted to give a context for the area under the curve value of 0.86; they could have pointed out, for example, that a perfect test has value 1 and random guessing 0.5.

Question 6
A minority of candidates attempted this question on case-control studies. A few of them gave excellent answers, but others were very poor. Again, some candidates struggled to write short notes for parts (i) and (ii). Several candidates wrongly used the table of data presented in the question in its given form as the basis for an unmatched analysis (such as a chi-squared test of association); they lost a lot of marks as a result.

Question 7
This question examined the theory of simple random sampling from a finite population. Most candidates answered it, gaining about half marks for it on average. Part (iii) caused particular problems, apparently because a substantial number of candidates were not familiar with the required proof.

Question 8
This second question on sampling was about stratified random sampling. About half of all candidates attempted it and they got about half marks on average. Several candidates failed to account correctly for the differential costs of sampling in different strata when constructing an optimal allocation scheme.