ROYAL STATISTICAL SOCIETY
DATA | EVIDENCE | DECISIONS

ROYAL STATISTICAL SOCIETY EXAMINATIONS, 2016

REPORTS OF EXAMINERS

GENERAL COMMENTS

This report incorporates the comments made by examiners after marking the papers set in 2016 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 2017 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 the light of the Royal Statistical Society's decision to withdraw from offering its professional examinations after 2017, candidates, and those advising them, should consider very carefully the number of modules for which they chose to register in 2017, particularly at the Higher Certificate or Graduate Diploma level. Since this will be the last chance to take the modules, there will be a temptation to register for all the remaining modules required to complete the awards. But, against this, candidates are most likely to be successful in passing if they are realistic about the amount of time they have available for study and enter for an appropriate number of modules. For candidates who are part-way through a qualification after the May 2017 examinations, the Society has undertaken to provide individual guidance on transitional routes to a qualification via one of its accredited partner institutions.

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 2016 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 2016 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. 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 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.) 4

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)

Overall, candidates performed well on this examination paper. The vast majority of candidates answered all eight questions.

Question 1

Candidates answered this question well and demonstrated an understanding of how they might perform simple random sampling and systematic sampling. However, very few candidates attempted to give a definition of simple random sampling in part (i) and no candidates gave an entirely correct definition.

Question 2

The majority of candidates performed extremely well, many achieving full or almost full marks for demonstrating an excellent understanding of the advantages and disadvantages of different data collection methods.

Question 3

Most candidates answered this question very well, although some candidates showed a very limited understanding of the differences between primary and secondary data.

Question 4

On the whole, this question was answered very well.

Question 5

This question was answered reasonably well. Most students designed the closed questions well although some questions could have been clearer and candidates should take care to specify units of measurement and time periods. Part (iv) was answered very well.

Question 6

The majority of candidates demonstrated an understanding of the differences between an experimental study and an observational study although some candidates appear to confuse an observational study design with *observation* as a data collection method.

Question 7

Generally, this question was answered well especially part (iii) on reasons for carrying out a pilot survey. Most candidates did not answer part (ii) on non-response bias in enough detail.

Question 8

This question was answered reasonably well and most candidates demonstrated an understanding of a large survey and sources of error. Candidates should ensure that the particular errors discussed relate to their chosen survey.

Module 2 (Analysis and presentation of data)

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

Some common problems remain.

- Many candidates fail to answer exactly the question asked, and so lose some or all of the marks available. A particular issue is that some candidates seem unwilling or unable to draw statistical diagrams even when these are specifically required by the question.

- Candidates' presentation is often poor. Illegible handwriting will inevitably 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

Candidates generally showed a good appreciation of the faults in the newspaper survey. However, some of the remedies suggested were inappropriate or unfeasible. For example, a newspaper is most unlikely to work with banks to generate unique IDs for respondents to a survey.

Question 2

Most candidates were able to extract the core messages from the data in this question. The fact that the question was heavily structured no doubt helped to keep candidates on track.

Question 3

Calculating the mean from a frequency distribution should surely be routine for these candidates. However, there were many errors. Most commonly, candidates calculated the mean of the frequencies.

In part (iii), candidates were asked to write a report on the data suitable for a radio audience. The best attempts were interesting and inventive. The worst were completely inappropriate – simply listing the figures, for example.

Question 4

The first two parts of this question, the diagram and the calculation, were done well. The interpretation in parts (iii) and (iv) was much less successful. There is no very clear relationship between the two variables; perhaps candidates were not comfortable saying this.

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 question was answered well by many candidates. The only common problem was in scaling up the figures in part (ii). Quite a few candidates simply added the same number to each figure to give a total of 100.

Question 7

The calculation of the mean and standard deviation in part (i) should have been easy, but some candidates got into a mess. Common errors included using the frequencies as the data, and confusing standard deviation and variance. The calculations of the mode, median and quartiles were done better by most, but not all, candidates.

Most candidates gave a correct explanation in part (iv).

Question 8

In this question, parts (i), (ii) and (iv) were routine and generally done well. The requirement in part (iii) to calculate the average annual rate of inflation defeated most, however. The usual approach was to divide the 10-year inflation figure by 10.

Question 9

This was found to be quite straightforward by most candidates. The trend and variation was clearly described in part (i). In part (ii), however, there was a curious and common misinterpretation to the effect that the diagram had a new data point added at zero on the vertical axis. It should have been clear that would make no sense.

The calculations in part (iii) were generally done successfully, though it was noticeable that many candidates took up a lot of space here. The predictions in part (iv) were generally done correctly and the comments on extrapolation were usually correct.

HIGHER CERTIFICATE

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.

An important element in this paper – indeed in interpreting statistical data generally – is the ability to extract a clear message from a graph or a table of figures. Weaker candidates often seemed unable to do this, preferring instead to repeat the data in their answer.

Question 1

This was the least popular of the four questions, perhaps because it was the least structured question or perhaps because there was so much data to absorb.

As usual, weaker candidates tended to regurgitate the figures in the table rather than interpret them.

The question asks very specifically for comparisons within and between groups of countries. This requirement was frequently ignored in favour of analyses of individual countries.

The question also asks for appropriate diagrams. However, most candidates drew no diagrams at all. The diagrams that were drawn were often of poor quality.

Question 2

This was a popular question and it was clear that most candidates were well informed on the nature and purpose of a census.

Many of the answers were thoughtful and interesting, particularly in respect of part (iii) on the claim that 'Jedi Knight' is a religion.

Question 3

In this question candidates were asked to comment critically on two published statistical diagrams.

The first diagram had several weaknesses that most candidates did not notice. For example, the column widths don't represent anything when they could represent the relative sizes of the races. Also, the diagram looks like a stacked bar chart, but it isn't. When asked what additional data would be useful, candidates tended to go for factors that the survey was not considering such as age and gender. More obvious factors such as relative numbers in the various groupings were usually overlooked.

The second diagram was discussed rather more successfully – perhaps because it has fewer faults than the first diagram. The message in the data was generally well described.

Question 4

The first part of this question was – or should have been – very straightforward as it involved fairly routine calculations. However, converting standard errors to standard deviations defeated many, and it was common not to find the *weighted* mean when combining different sized samples of men and women.

In part (ii), almost everyone commented that milk consumption rose from Q1 to Q4. Only a few, however, noted that the rise is massive.

In part (iii), the pattern of rice consumption was described, but it was rare for anyone to comment that it is the only food type to show that pattern.

Part (iv) was well done.

Various approaches were used in part (v) with varying degrees of success. It was not sufficient simply to comment on the total food consumption of men and women. The question required comments on the main food types.

Module 2 (Probability models)

In general some of the candidates did well on the paper, where they showed a good understanding of the normal distribution results and using the continuous random variable density and cumulative distribution function. This was reflected in the last two questions on the paper being answered well.

Question 1

Some candidates struggled with the ability to prove results. In part (i) it was a very straightforward algebraic manipulation to prove the result, but proved to be difficult for some. It is important that candidates can derive mean and variances for the distributions. In general there was a lack of confidence in the algebraic manipulation to perform the proofs required in this question. In part (iv) this combined the combination of Poisson random variables and the use of conditional probability. Some candidates did not see that belts C and D were the only options and hence can use the conditional probability formulae that they know. In (iv) (b) this required combining Poisson rates from belts C and D and extending that for a 5 day period. This proved difficult for some.

Question 2

This question required writing down the pmf and finding the expectation for the number of failures before the first success. This was not always done well. It is important that a candidate understands how to derive a pmf and can then use it to derive means and variances etc. In part (iii) they needed to apply similar ideas to a practical problem. In part (iii)(a) some candidates ignored the probability of the success. Part (iv) was done very easily if you used the results of part (ii) by running the experiment twice. Only a few candidates applied this result. It is important to keep in mind what they have done earlier in the question.

Question 3

This question was in general done very well, where candidates were able to write down the correct linear combination of linear random variables in part (a). In part (b) the central limit theorem was required and knowledge of the distribution of the sample mean. This was done well and the need to find a sample size if important.

Question 4

This was also done well in general. It involved finding the cdf and probabilities using the latter. Some candidates struggled to find the median. The final part gave anther density and needed calculation of the mean and a probability.

Overall greater understanding of deriving discrete densities and more confidence/understanding proofs is important. Candidates need confidence in algebraic manipulation so that they can attempt questions.

Module 3 (Basic statistical methods)

As a general comment, it was good to see that candidates attempted just three questions rather than trying all four this year, thereby giving more thinking time for their chosen selection. A further, but rather more negative, general comment is that candidates need to ensure that they read the questions well and answer exactly what is asked, as time can be wasted writing down irrelevant information and gaining no marks.

Question 1

The definitions in Question 1 should be well known, and candidates needed both to relate them to the probabilities of the Type I and Type II errors as required AND also to give definitions. For example, it was not relevant to the question to state that common significance levels are 5%, 1%, etc as this is not part of a definition. In (ii), parts (a) and (b) were generally well done. However, (iii) and (iv) both required answers which were probabilities whereas many candidates performed tests which were not relevant and were not awarded marks.

Question 2

The model required in Question 2(i) was the Binomial with $n = 10$ and $p = 0.12$, although many candidates quoted the Poisson distribution. When specifying a distribution it is important to give the parameter values in addition to the name of the distribution. Part (ii) was generally well done, requiring pooling of categories to ensure that all expected frequencies were large enough. It is not acceptable to state the null hypothesis as “the expected and observed frequencies are close” (which could relate to ANY goodness of fit question) but the hypotheses should be related to the actual situation. Part (iii) required looking at the direction of the discrepancies between the given observed and expected frequencies, which led to the conclusion that the Stalwart seeds were more successful in germinating than previous varieties.

Question 3

This was the most popular question and was generally well done. The confidence interval in (i) was constructed using a t value, not a z value, as the sample size was small and the population variance unknown. (ii) was a paired test, the strict assumption being that the single sample of differences has been drawn from an underlying Normal distribution. In (iii) candidates needed to be clear which Wilcoxon test they were recommending (Signed Ranks or Rank Sum).

Question 4

In parts (i) and (ii), the confidence intervals were constructed using t values rather than z values again, this time on 14 degrees of freedom. Parts (iii) and (iv) were standard applications of the variance ratio F test and the two-sample t test. When calculating intervals or performing tests it is always important to mention the appropriate degrees of freedom and any tabulated values used.

Module 4 (Linear models)

Question 1

This question was well done. Some candidates did not know how to calculate an estimate of the variance, so were unable to calculate the variance of the slope parameter.

Question 2

Again this question was very well answered. Some candidates are unaware of the tables given for testing the sample product moment correlation coefficient. In part (b) some students ignored re-ranking the scores.

Question 3

Some candidates were unaware of the correct model. Some candidates put in regression slope parameter when there are no explanatory variables. Most candidates came up with good ideas on the problems with the paint ship design of the experiment.

Question 4

Candidates need to ensure they understand the difference in testing the overall regression and individual coefficients.

Module 5 (Further probability and inference)

As a general comment, it was good to see that candidates attempted just three questions rather than trying all four this year, thereby giving more thinking time for their chosen selection.

Question 1

This question relied on the geometric distribution, although it was not necessary to know this in order to do the question. There were relatively few completely correct answers to (i) which required summation of a geometric progression. However, derivation of the likelihood function in (ii) was much more successful. Those who took the logarithm before differentiating were considerably more successful in finding the MLE and demonstrating that the likelihood was actually a maximum than those who proceeded directly with the likelihood itself. Inability to find the second derivative also hampered calculation of the confidence interval in (iii).

Question 2

Part (a) was well done by those who could successfully differentiate a logarithm. Demonstrating the results in (ii) relied on Mx(0)=1, which needed to be justified rather than assumed without explanation. In (b) the proof of the Poisson mgf was generally well done, although candidates need to explain clearly how the exponential expansion is used. In (ii) the expressions for the expectations of Y squared and Y cubed were given on the paper, but quite a number of candidates decided to demonstrate these results from the mgf. This wasted valuable time and gained no marks. (iii) was done by only a small proportion of those who attempted the question, although the required distribution (Poisson(nxlambda)) seemed to be quite well known.

Question 3

This question was very popular and generally very well done. In (iii) the conditional pmf was best presented as a table or list of possible X values with their corresponding probabilities. Candidates need to ensure that they present this in full when asked for it. The calculation of the covariance in (iv) required E(X) and E(Y) obtained from the corresponding marginal pmfs. These should also be presented in full so that it is clear where calculated values have come from.

Question 4

This was the least popular question. Although the proof in (i) is a well known result, it required careful explanation and manipulation involving summation and expectations. In (iii) the expression for the MSE needed to be minimised with respect to *a* which involved differentiation with respect to *a.* Some candidates expanded out the squared bracket term which made the differentiation and ensuing simplification much more difficult, it was easier to use the chain rule to differentiate the squared bracket whence the result could be worked out in just a couple of lines.

Module 6 (Further applications of statistics)

No detailed comments

Module 7 (Times series and index numbers)

The results on this paper were encouraging in that the majority of candidates showed their knowledge in both subject areas and completed the required calculations accurately. The best papers demonstrated real excellence.

Question 1

This question covered two methods of forecasting a time series with trend plus quarterly seasonal effects. Here several candidates did not realise that the seasonal components should sum to zero. The second model was a multiple regression on time and a 4^{th} quarter effect. Here candidates needed to use both the estimate and standard error, either via a t-ratio or confidence interval, to judge whether the variables should be retained.

Question 2

This question was on ARIMA modelling of a weekly data set. Although mostly well-done there were some misunderstandings.

Firstly not every time series has a seasonal pattern and here it was a mistake to imagine cycles in random noise. Secondly very few candidates understood that the fitted constant in the model for the first difference of the logged data corresponded to the slope in the logged series itself. Finally it was disappointing to see some candidates trying to interpret the lag zero residual autocorrelation since this is, of course, always one by definition.

Question 3

This question was on the relation between the Laspeyres volume and Paasche price indices and the use of the latter as a deflator. In the numerical application most candidates, but not all, grasped that they should apply the deflator separately to each product group before combining them.

Question 4

Part (a) concerned the representation of the Laspeyres price index as a weighted average with an extension to two groups of commodities. The latter required an understanding of the reason why weighted averages are needed ie that the indices cannot simply be added.

Part (b) was a numerical calculation of the Laspeyres index plus chain-linking. Again this was for the most part done competently though a few candidates seem to have confused chain-linking with rebasing.

Module 8 (Survey sampling and estimation)

Candidates are reminded that they are expected to memorise certain formulae as indicated in the syllabus, and know how to apply these formulae to data.

Question 1

This question on sample size for stratified random sampling was attempted by half of all candidates. Several candidates achieved very good marks. Some candidates lost marks by not completing all parts of the question.

The earlier parts of the question introduced the formula for the sample size to be used in part (iv). In defining the terms, some candidates thought W_h was the number of units in stratum *h.* Please note W_h is the population stratum weight in stratum *h*, defined as $\left(\frac{N_h}{N}\right)$, and N_h is the number of units in that stratum. Those who attempted part (ii) did it well, writing down the formula for a 95% confidence interval and re-arranging terms to obtain an express for V and d, which was pleasing.

Most candidates were able to apply the formula in part (i) to calculate the sample size in part (iv). However, there were a lot of mistakes some attributed to the definition of W_h .

Question 2

In this question, candidates had to write short notes on how selection bias might relate to each of three methods used by polling organisation. There were some good answers but sometimes candidates did not give enough relevant points to obtain full marks (or wrote extensively on one aspect). Many candidates lost marks in discussing the merits of each method and not the contrasting these methods as required. For this referendum, 16 and 17 year olds were able to vote. Only one candidate mentioned the difficulties of identifying people in this age group.

In part (ii), candidates were asked to suggest a possible question to find out about people's voting intention for the referendum. Many candidates' question favoured one outcome; some candidates misunderstood the question and asked if people were going to vote on the day of the referendum.

Question 3

This question presented data for analysis on library usage. There were good marks to all parts of question 3. However, it was noted that sometimes candidates were rounding too early in calculations.

Candidates were asked to obtain a point estimate and approximate 95% confidence intervals for the proportion of students attending the library and the total number of visits. A common mistake was to estimate the population mean rather than total. Some candidates attempted to estimate the proportion of students visiting a public library using the sampling fracture, $\frac{n}{N}$. The data were presented as a frequency table and a few candidates struggled to calculate $\sum x$ and $\sum x^2$.

In part (b), there was a lot of discussion on whether systematic sampling should be treated as random or non-random. Half of the candidates thought it was random in that the starting point is random. It is strictly non-random.

Question 4

This question on ratio estimation and the use of cluster sampling was attempted by two-thirds of all candidates. Several candidates achieved very good marks.

Most candidates were familiar with the concept of correlation and how it relates to the ratio estimator. However, the discussion of properties of the estimators (in terms of bias and precision) was often omitted (or misunderstood).

Part (ii) (b) was well done. However, it was noted that some candidates were not using the value of the estimated standard error of the ratio estimator of the population total as given. Candidates were inappropriately calculating a standard error using the formula for the mean of a simple random sample, using the value (8550.947) as the standard deviation.

Candidates' discussion of the practical difficulties that might arise in carrying out cluster sampling was disappointing. Not everyone understood the definition of a cluster and how it differed from a stratum, and the practical challenges in defining clusters.

GRADUATE DIPLOMA

Module 1 (Probability distributions)

It was a year of mixed fortunes for the candidates who sat this paper, with several producing excellent performances but others scoring very low marks. A good part of the explanation for this variety of outcomes is that just three-quarters of candidates were able to start five questions.

Question 1

Almost all candidates attempted this question on discrete random variables in one and two dimensions. They scored under half marks on average. Candidates generally did well in part (a) but less well in part (b). They encountered a variety of problems with part (b) (i) which they then carried over to (ii): most attempts did not specify the range space of S or note when they were relying on the independence of X and Y; a large number forgot to sum over x from 1 to s-1; several failed to realise that the final sum was just (s – 1) times an expression in θ.

Question 2

The vast majority of candidates tackled this question on the gamma distribution and an extension of it into two dimensions. The average mark for this question was higher than for other questions on the paper. Most candidates were able to write down the standard proof required in part (i) without too much difficulty. In parts (ii) (a) and (b), the biggest problem was that candidates failed to specify the correct limits of integration given that the joint range space was not rectangular. Several would have saved themselves considerable amounts of time by following the hint on the question paper and using the results from part (i) to obtain the expected value and variance of X. A surprisingly large number of candidates seemed not to know the law of iterated expectation, which was required for (ii) (c).

Question 3

Only a small number of candidates tried Question 3 on survival and hazard functions; some attempts were excellent (and scored full marks) while others were really poor. Generally, attempts at part (i) were much better than attempts at (ii) and (iii) – which might be due to the fact that (i) is a standard proof.

Question 4

This question, on the multivariate normal, was another that was not very popular with candidates. Most of those attempted it, however, scored good to excellent marks. A common error in part (a) was to miss out or use an incorrect constant when calculating the expected value of Y_1. In part (b), candidates generally got the correct expressions for expectations but many did not derive the correct variance or co-variances.

Question 5

This question examined work on order statistics and bivariate distributions. About 40% of candidates tried it, most of them very successfully. In part (i), a number of people failed to substitute the correct functions, $f(x) = 1$, $F(x) = x$ $(0 < x < 1)$, into the given formula. Attempts at part (ii) were very good, though not everyone was able to perform the integration correctly. Some candidates struggled to relate part (iii) to results obtained in the earlier parts of the question, often because they failed to appreciate that θ was a constant.

Question 6

The vast majority of candidates attempted this standard question on moment-generating functions and the Central Limit Theorem, in the context of the normal approximation to the binomial distribution. Their average mark was good though a surprising number of candidates struggled with all parts of the question, even the standard bookwork of part (i).

Question 7

This question was based on the simulation part of the syllabus. About half of all candidates tried it but few of their attempts were successful. Candidates could have saved time by realising that the Poisson tables provided gave the values of the cumulative distribution function required for part (i). There were a lot of mistakes in finding the integral in part (ii). Few people even attempted part (iii).

Question 8

About half of the candidates attempted this question, on bivariate continuous distributions. Their attempts were either very good or rather poor. Part (i) was generally done well, parts (ii) and (iii) less so.

Module 2 (Statistical inference)

Performance on this paper was very variable. There were some excellent candidates, but a disappointingly large number of candidates who were ill-prepared and scored very few marks. The first four questions were the most popular. I suspect that some candidates would have done better if they had attempted one or more of the later questions. It is always a good idea to read the whole paper before deciding on choice of questions.

Question 1

In parts (i), (iii), (v) there was a worryingly widespread confusion between sample and population quantities. Parameters, estimates of them, and expectations were frequently mixed up.

In part (v) many candidates chose to evaluate mean square error as variance + bias. Whilst correct, it was easier to use the basic definition given in the question. Some candidates also made things more difficult for themselves by not using the expectations given in the question.

When minimising MSE it is necessary to check the second derivative to see whether you have a maximum or minimum. A similar comment applies when finding maximum likelihood estimators in Questions 2 and 4.

Question 2

The most popular question but the second least well done. Most candidates made the binomial likelihood in part (ii) much more complicated than it need be and very few successfully attempted part (iii).

Question 3

Very few candidates knew the definition of sufficiency. Most avoided part (v).

Question 4

A majority of candidates gave a likelihood ratio test, rather than a GENERALISED likelihood ratio test in part (i).

Question 5

No specific comments.

Question 6

This was the least popular question, though there were still 15 attempts. In maximising the absolute difference between the null and empirical cdfs (the test statistic) in part (ii), everyone evaluated the difference only at the change points and did not recognise that it was larger just below the change points.
In part (iv) no-one recognised that the test statistic was the number of observations less than the MEDIAN (not mean).

Question 7

The best answered question.

Question 8

There were relatively few attempts (possibly because of its physical length), mainly by candidates who had done badly on other questions and repeated that performance, making it the least well done. Only one candidate succeeded with part (iv).

Module 3 (Stochastic processes and time series)

Results this year covered almost the full range of marks and the quality of the best attempts was outstanding. However there were also candidates who were clearly not sufficiently prepared notably some who answered only three questions.

This year the eight questions were equally split between Stochastic Processes and Time Series.

Question 1

This question on a Markov chain in a health-care context was both the most popular and the highest scoring. Most candidates were able to calculate and interpret the relevant conditional probabilities and the stationary distribution.

Question 2

The second question, on branching processes, was attempted by two-thirds of the candidates. The initial parts asked for explanations of the initial steps in the proof in part (iii) and provided some notation to help. The marks here were very mixed. Part (vi) asked for the numerical calculation of the extinction probability using the equation derived in (i) – (v). There were many accurate attempts although a few ignored the final hint.

Question 3

In this question on a queueing system with state dependent arrival and service rates, the general formula for the equilibrium probabilities was given. Candidates who derived these were therefore wasting valuable time. Several candidates clearly had a good understanding of this topic and there were some complete or near-complete answers.

Question 4

This question developed theory for the Poisson process with a time-varying rate. Most candidates who attempted this question could provide or reproduce the derivations though perhaps missing some steps or explanations.

Question 5

Surprisingly the first of the time series questions was the least popular overall. Part (a) was a straightforward piece of book-work on simple exponential smoothing followed by a numerical example. Part (b) asked for the calculation of the autocorrelations for a simple ‘state-space’ model based on two independent white noise series. This may have been unfamiliar but is actually a fairly straightforward test of understanding.

Question 6

The most popular time series question concerned properties of a second-order moving average process, its forecasts several steps ahead and the corresponding error variances. The latter topic was a weak point for some candidates.

Question 7

Although questions about the characteristic polynomials of ARIMA models are a time series staple, this turned out to be the lowest scoring question. This may have been due to unfamiliarity with the application which concerned the problems created by trying to fit a redundant model to data. A few candidates confused stationarity and invertibility.

Question 8

Many of the candidates clearly welcomed this data-driven example on identifying, fitting and checking ARIMA models and it had the second highest average mark. However a surprising number of answers to part (iii) based their argument either only on the estimates or only on the standard errors from the computer output instead of combining them into t-statistics or confidence intervals.

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 important 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, taking account of the context of the analysis.

Question 1

On average the best answered question, though there were some very good answers, and others that displayed a complete lack of understanding of these analysis methods. Most answers correctly identified the least-squares estimator and stated the Gauss-Markov theorem, though some described the estimator as "best" rather than as having minimum variance, and some forget about the estimator being unbiased. There is clearly some confusion about multicollinearity, with some answers just assuming that this meant linear associations between 3 or more explanatory variables, and not also pairwise linear correlations (though this is sometimes referred to as collinearity). Descriptions of stepwise variable selection approaches were generally good, though many answers failed to identify the influence of the choice of variance ratio thresholds for selection or elimination of variables. Most answers failed to interpret the final selected regression model, with some failing to identify multicollinearity as a cause for *Ht* being omitted. Descriptions of the underlying assumptions were generally clear, though often omitted the assumption about linearity of the relationship.

Question 2

Some very good and complete answers, but also many answers with important elements missing. There was some uncertainty about the orthogonal nature of the design, with a number of answers confusing the blocking effects of row and columns with the treatment effect. Most answers demonstrated an understanding of the calculations of sums of squares, though there were a number of careless errors, including of degrees of freedom, which led to incorrect mean squares and F-values, and hence incorrect interpretations. Some candidates failed to interpret the importance of blocking effects, though there was one good illustration of the gain in precision through blocking. Many candidates were not clear on the use of orthogonal contrasts to partition sums of squares, though some calculated and interpreted the polynomial contrasts well, if failing to test for the lack-of-fit of the quadratic model. There were some careless errors with calculating confidence intervals for the difference between treatment means, such as being based on totals, using the Normal distribution, or using the standard deviation rather than the SED.

Question 3

The early parts of this question were reasonably straightforward, but this was, on average, the worst answered question, primarily because no candidates were able to construct the allocation of treatments to blocks. Possibly this reflects that the cyclic approach to the construction is not commonly used or really understood, but many candidates failed to even show the start of an approach to this part, even given the hint provided. In general candidates were not clear on the relative advantages and disadvantages of RCBDs and IBDs for the presented scenario – the key is in considering the effect of the constraint of the number of units per processing batch being fewer than the number of treatments on within-block variation in the complete book alternative, and the potential loss of balance with the incomplete block alternative. Most candidates successfully identified block sizes and numbers for the different design parameters, though many could not explain the key relationship relating to the total number of paired comparisons for a given treatment. A description of the randomisation process was often not attempted after failing to develop a treatment allocation, but should have considered the allocation of sets of blocks to days (and within days), the ordering of treatments within blocks, and the allocation of treatment codes to treatments.

Question 4

Most candidates who attempted this question obtained good marks for the first two parts, but lost marks with the interpretation of the analysis and extension to consider lack-of-fit. The key point in explaining the difference between linear and non-linear models is identifying that the response needed to be linearly related to all of the parameters. There were some very clear explanations of the Newton-Raphson procedure, though often with far more algebraic detail than was justified or expected given the number of marks available. There was some confusion in identifying whether models in part (ii) included non-linear parameters, with some candidates failing to classify parameters as linear or non-linear, and many candidates failing to identify the potential linearization of model (d) through fixing the value of *w*. The interpretation of the analysis results was rather mixed, with many answers fialling to describe how the individual model parameters related to the sigmoidal shape of the response, and some confusion about how parameter *b* related to the maximum slope. Most answers correctly inverted the equation to express *x* in terms of *y*, though there were a few careless calculation errors. A few candidates provided very clear explanations of how the replicate values at each time could be used to obtain a measure of the pure error and hence a test of the lack-of-fit, though some answers failed to explicitly identify that it was the replicated *x*-values that allowed this extension.

Question 5

A very popular question but generally answered rather badly, usually because of a poor interpretation of the analysis of variance table and tables of means, or because some parts were not attempted. Many interpretations of the ANOVA table just identified the significant model terms, though some missed the Manufacturer. Temperature interaction and many failed to describe the patterns of response to the different (combinations of) factors. Some answers sensibly started by considering the significant interaction effects, though often then just described the main effects within these interactions rather than identifying the cause of the interactions. But some answers completely ignored the tables of interaction means.

Many answers also struggled to present the conclusions in non-technical language – this shouldn't mean failing to comment on the statistics significance or trying to explain the concept of an interaction, but providing an interpretation of the (combined) effects of the factors taking account of the underlying scientific context. Generally, candidates provided a clear identification of the underlying assumptions, though there was some confusion with the additional assumptions required for regression modelling, and many answers failed to describe how the residual plots were constructed (in terms of how the values to be plotted were obtained from the data).

Relatively few answers identified the binomial nature of the percentage data being analysed, and hence the expectation that the variance would vary with the mean. Those that did identify this characteristic generally identified this pattern in the fitted values plot, though some answers focussed more on lack of Normality – one good answer identified the histogram as being more peaked as a result of varying levels of variance at different mean levels. A few answers successfully identified the use of an appropriate transformation (logit, arcsine) to cope with the heterogeneity of variance, with a number also identifying the use of a GLM (though with some confusion about the appropriate distribution and link function).

Question 6

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 design, but often failed to discriminate between defining (fractionating) contrasts and confounding contrasts, and to identify the impacts of the choice of these contrasts on estimating high-order interaction terms and on the aliasing structure. There was some confusion about using these contrasts in constructing the design – it is important fractionate before confounding with blocks. Where sensible fractions were defined, good attempts were made at allocating treatment combinations to blocks – but some answers failed to generate the half-replicate first. Most answers also failed to consider the pattern of two-factor interactions that needed to be estimable. The outline ANOVA table needed to identify those terms that were confounded, and that all important 2-factor interactions could be estimated. Where the outline ANOVA table was constructed, generally the assumption about high-order terms being allocated to error was correctly identified.

Question 7

A popular question, though with a rather variable quality of answer, with some very poor answers. Many answers struggled to discriminate between fixed and random effects, often being confused by thinking about fixed and rand models, though were often then able to identify the fixed and random terms in the example. The identification of clear examples appeared to help with this discrimination. Some answers struggled with the hierarchical structure of the example, failing to identify locations nested within types and occasions nested within locations, even though these were suggested in the partially completed analysis of variance table in part (iii). Many answers correctly identified the terms in part (ii) but did not clearly state the assumptions associated with each term for the analysis. There were numerous errors in identifying the degrees of freedom for the “locations within types” term, usually considering location as being crossed within type rather than nested, and this led to incorrect numerical values and to incorrect interpretations. Some interpretations also failed to take account of the hierarchical structure in specifying tests, and failed to specify the interesting hypotheses being tested, though generally showed the appropriate testing approach. Most answers failed to discuss the relative sizes of the variance components (between occasions, between locations).

Question 8

Not a particularly popular question, with both some very good and some rather poor answers. Most answers correctly identified the components of a GLM (exponential family of distributions, link function, explanatory model) though a few failed to show how the response was related to the explanatory model. Most answers also correctly identified the example as being a Poisson GLM with a log link function. There were some good attempts to describe iteratively reweighted least squares, though many answers struggled to identify the reweighting element. Interpretation of the output was rather variable, with many answers just focussed on the analysis of deviance table and some lack of clarity about deviance ratios and the associated p-values. A few answers attempted to identify the efficacy of the treatments, though most only managed to do so for the tachyporus effect. Many answers failed to describe the concept of over-dispersion, often being confused with the deviance ratios for treatment term, though some did identify the estimated dispersion parameter in the output.

Module 5 (Topics in applied statistics)

Candidates generally performed below average in this paper although several achieved excellent marks in it. Some candidates seemed completely unprepared for this paper and scored low marks for every question attempted. Candidates mainly attempted questions 3 to 8 as the epidemiology/demographic questions were not popular and the answers given poor. By neglecting part of the syllabus candidates may reduce their chances of passing. Good marks were generally achieved when candidates answered all parts of a question. Many candidates did better when carrying out calculations than answering the more explanatory questions or interpreting results. Often the practical context was ignored in answers and candidates should practice writing up analyses which utilise the real-life context in their answers.

Question 1

Not a popular question on epidemiological methods with poor answers, although one candidate achieved excellent marks in it. In part (i) most candidates could not express these concepts in terms of conditional probabilities and often there was confusion on what event is conditioned on in a cohort and case-control study. Few candidates could prove that the odds ratio for a case-control study equals the odds ratio for a cohort study. Most marks were obtained for parts (iii) and (iv).

Question 2

Not a popular question on epidemiological methods with poor answers. Many candidates couldn’t correctly define the standardised mortality ratio and couldn’t describe indirect standardisation in words. In part (iv), almost no-one could relate the Poisson regression model to the concept of indirect standardisation.

Question 3

About 80% of the candidates tried this question on principal component analysis (PCA) and they gained just under half marks for it on average. In part (i) many couldn’t list and discuss the decisions that need to be made when carrying out a PCA even though these decisions were implicit in answering the other parts of the question. In part (iii) many didn’t state the obvious that all the correlations were positive and the sizes were moderate. In part (iv) some candidates lost marks because they gave a mathematical interpretation rather than in the substantive context, e.g. that good students tend to do well in all subjects and bad students tend to do badly in all subjects. In part (v) the criteria used to decide on the apparent dimensionality of these data were not explicitly stated and marks lost.

Question 4

About 75% of the candidates tried this question on cluster analysis and they gained just under half marks for it on average. In part (i) many candidates had good answers, but the decision on whether to transform the data, e.g. by standardising the variables or whether to use the raw data and the decision on choosing the final number of clusters by some criteria were omitted. In part (ii) many candidates didn't recognise that the main difference between hierarchical and non-hierarchical clustering was that the user needs to choose the number of clusters in hierarchical clustering, but in non-hierarchical clustering it is pre-specified. In part (iii) many candidates didn't state that the visual inspection of a dendrogram was useful to recognise distinct clusters. In part (iv), while most candidates drew the dendrogram correctly, a few didn't calculate the correct distances. In part (v), reasonable answers were given from those who used the correct dendrogram.

Question 5

About 90% of the candidates tried this question on survival analysis and on average this was the best answered question. In part (i) a sizeable minority didn't know these definitions, which caused difficulty in answering part (ii). In part (iii), the explanations of what is meant by right-censored observations were weak. In part (iv) plots and calculations were sometimes done for both groups combined rather than separately. Often the Kaplan-Meier calculations were done incorrectly for the experimental group with censoring. Sometimes the Kaplan-Meier graph was plotted as a line graph rather than a step function. In part (v) only a few candidates made the connection that for the control group with no censoring that the hazard rate estimate was the inverse of the mean survival time.

Question 6

About 75% of the candidates tried this question on survival analysis and they gained just under half marks for it on average. In part (i) most candidates answered correctly, with the exception of some candidates didn't understand what a cumulative hazard is. In part (ii) most candidates did well, but few could not write down an expression for the hazard ratio for this model. In general, most answered well for part (iii). In part (iv) many didn't note that the exponential distribution is a special case of the Weibull distribution and use the 95% confidence interval calculated in part (iii) to answer the question. In part (v) many didn't realise that the plots should be both straight and parallel to justify the Weibull proportional hazards model.

Question 7

About 70% of the candidates tried this question on survey sampling and most gained under half the available marks. In part (i) many didn't define the finite population variance in symbols. Answers to part (ii) were very poor as many candidates didn't know the formula for the finite population variance. In part (iii) most candidates couldn't correctly calculate the finite population variance and some did the confidence interval for the mean, not the total number of office staff. In part (iv) many did not know how to start to answer this question and to assume the worst case scenario of the percentage of one-half for the calculation. As the achieved sample size implied that only 16 members would not be sampled, the best advice would be to use a census rather than a sample. In part (v) reasonable answers were given when attempted.

Question 8

About 60% of the candidates tried this question on survey sampling and most gained under half the available marks. In part (i) the definition given was often imprecise. In part (ii) the examples given were wrong or not very good. In part (iii) what was being estimated, i.e., the mean, was often omitted and answers to why these allocations are only theoretical was very poor. In part (iv) many didn't know the optimal allocation formula and when stated, many didn't realise that the formula involves standard deviations, not the variance.