ROYAL STATISTICAL SOCIETY EXAMINATIONS, 2005

REPORTS OF EXAMINERS

General comments

Most comments made by examiners refer to specific features of questions set in this year's papers. But every year examiners draw attention to aspects of examination technique which could be improved. As we have noted in earlier reports, it is disappointing to see candidates losing marks unnecessarily. Several comments made by examiners in 2005 echo those made in recent years. We therefore repeat here the advice given at the start of last year's report, revised so as to incorporate further general comments made by examiners following the 2005 papers.

- Read any question you intend to answer slowly and carefully, and ensure you answer the question actually asked. Every year, some candidates reproduce bookwork which may have some relation to the topic, but does not answer the question itself. Examiners award marks in accordance with detailed marking schemes, which assign marks for specific answers to each part of each question. There is therefore no point in writing down what you know about a different (if similar) topic, since the marking scheme will have no marks available for this.

- Make sure you understand the difference between the instructions *explain* and *define*. An *explanation* of some concept requires one or more sentences, in which the concept concerned is described, and (if appropriate) the purpose or use is outlined. In a mathematical examination, a *definition* is a short and precise statement, which may require the use of mathematical notation. If a definition is required, a rough description is likely to be awarded no marks.

- Take note of the marking scheme printed on the paper. It is a waste of your time writing a detailed two-page description of some topic, if this can only be awarded 2 marks.

- 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 which does not pass through the main part of the data points is very unlikely to be correct.

- 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.

- If a rough sketch diagram is required, this can be done in your answer book; there is no need to draw it accurately on graph paper This might for example apply to a sketch of a probability density function. Of course, such sketches must always be sufficiently clear that salient features stand out properly. However, when an accurate graph or chart is required, this should always be done on graph paper; and you should make sure you include a title and label the axes. This might for example apply to histograms.

It is important to follow the instructions on the front cover of the answer book. We realise that candidates will not wish to spend time during the examination reading the front cover, so we have

produced a copy you can consult on the Society's website. You are strongly encouraged to look at this before the examination, and to ensure that you follow the instructions. 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.

4. Graph paper should be attached opposite the answer to which it relates.

5. Enter in the space below (NOT in the side panel) the numbers of the questions attempted.
     (Write these numbers *in the order in which you answered the questions.*)

Ordinary Certificate: Paper I

General

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

The overall standard was fairly good with some excellent answers from many candidates and a few candidates obtaining high total scores. There were also a few scripts of very poor quality where candidates appeared to know almost nothing about the topics on the syllabus. Candidates who did not make a full attempt at all the questions tended to fail overall.

Question 1

Unfortunately, this question was done very badly by all but a very small number of candidates. In part (i) hardly anyone described a two-stage sampling scheme (although multi-stage sampling is on the syllabus). A common misconception was that dividing population units into strata or into clusters could be considered to be the first stage of sampling, and that sampling from strata or taking a sample of clusters, using either simple random or systematic sampling, was the second stage of sampling. Some candidates appeared to think that in (one-stage) cluster sampling all clusters are sampled. There was also insufficient attention to detail as to what the units were at any stage; for example candidates wrote of putting schools into strata and of taking samples of ex-pupils from the strata. Some candidates described methods which did not result in a final sample comprising ex-pupils, as required.

In part (ii) many candidates wrote in general about the advantages and disadvantages of the four methods of sampling (simple random, stratified, systematic, and cluster) and therefore failed to relate their answers to the proposed survey at all, let alone to the answers that they had given in part (i).

Question 2

Most candidates did this fairly well, and a few produced questionnaires of a very high quality. Hardly anyone who gave options for occupation gave anything other than a paid job or study, although, for example, many school leavers do not enter a paid job (or proceed to further study) immediately. A few candidates lost marks because their questionnaires did not separate 2004 and 2005 occupations. It was pleasing that the majority of candidates gave an introduction to the questionnaire and details of how to return it, and added thanks at the end.

Question 3

Most candidates chose to design a database rather than a spreadsheet. This was done fairly well. Some candidates lost marks because they did not allow for storage of all the items specified in the question. Many candidates gave very long field names whereas shorter ones would have been better, and some suggested very long fields where it would have been better to have had several fields; for example, in the case of addresses. Many candidates did not appear to realise that answers to questions could usefully be coded before entry in a database; they simply left blank space to enter words (such as names of districts) or answers to open-ended questions asking about career plans.

Question 4

Part (ii) was done fairly well by most candidates, but part (i) was done relatively poorly. In part (i) some candidates commented on the size of sample and some on the fact that one region had been chosen from each of the two types of region, although what was required was a comment on the non-random nature of the sample. (A stratified random sample could be designed so that one region of each type was chosen.) Some wrote that there was no need to have a sampling frame, which was not relevant to the question. Others wrote at length about choosing merchants by a non-random method, although the question referred to the selection of regions.

There were many quite different but valid difficulties mentioned in part (ii), and there were interesting ideas as to how to address these difficulties.

Question 5

This question was well done by most candidates. In part (i) some candidates seemed to be confused as to whether the local representatives were the same as the merchants, and some appeared to have lost sight of the fact that it was background information that was required; that is, they did not realise that this part of the question was not asking about the survey of prices. Some were very concerned as to whether the person obtaining the information could speak English and wrote about translators, but there is no reason why researchers conversant in the language of the country should not be used.

In part (ii) most candidates missed the point that faulty memory would mean that it would be difficult to obtain accurate values for last year's prices, though other interesting points were made; for example, that there would be variability in prices both between sales outlets and even during a day at a single outlet. Some candidates wrote as if it would be the merchants who would be asked about last year's prices, whereas anyone in the country could be asked.

Question 6

The explanations of interviewer bias were poor. Stating that it is bias caused by the interviewer is not an explanation unless a definition of bias is given. Some candidates explained bias in terms of ways in which it could occur, but mentioned only interaction with the interviewee and omitted to say that interviewer bias could occur at the recording stage.

Examples related to the collection of prices were required, but some candidates gave general examples. Many candidates gave examples where the bias would have resulted from poor survey design; for example, where the interviewer was left to find a sample of merchants without quota controls or any other guidance, or where the interviewer was apparently not given a questionnaire but left to ask

questions, or perhaps was given a questionnaire with biased questions. These were not really examples of interviewer bias.

Question 7

Most candidates did part (i) correctly. On part (ii) candidates either completed it correctly or had little idea what to do, though most calculated $N_i s_i / \sqrt{c_i}$, which was not in itself worth many marks. Hardly anyone wrote the cost constraint as an inequality (using '$\leq$') but treated it as an equation. Some candidates worked out the costs using non-integer sample sizes. A few candidates lost marks because of a failure to give a clear statement of the conclusion.

Comments in part (iii) were mediocre for the most part. It should be remembered that to calculate an optimal allocation we use estimates of the cost and the SD for just one item (the staple vegetable), although, in practice, the survey organiser may well consider asking about other commodities.

Question 8

This question was not done well. Many candidates interpreted it as requiring a description of an official index number or a definition of a standard index number and then gave a fairly short comment that the price index for the country concerned could be found similarly (and some candidates made no comment about the country at all). The intention of the question was that the focus should be on a price index for fruit and vegetables for the country concerned. Few candidates mentioned that the price and weight in an index number would be an average of some kind, although many said that prices are usually collected from several outlets.

Ordinary Certificate: Paper II

General

The paper aims to test candidates' ability to perform basic statistical calculations and to draw charts. It is important that candidates can interpret their results.

It was pleasing to see that the vast majority of candidates attempted all eight questions. It was also gratifying that, on the whole, the standard of chart drawing was higher than in previous years, except in Question 4. All scripts were legible and the clarity of expression appeared to be an improvement on previous years.

There are certain basic formulae that need to be known. On this paper, there was evidence that those for standard deviation, rank correlation coefficient and price relative were not well-remembered.

Candidates should read the questions carefully and do exactly what is asked. This was not always the case for Questions 3 and 8

Question 1

This question was very well done with the basic rules of good chart drawing being followed. The comments were not always accurately stated. For example, the number of sufferers in the 75–84 age group in 2050 is expected to be double the number in 2000, i.e. an *increase* of 100% (not 200%) and the corresponding increase in the 85+ group is 344% (not 444%). The advantages and disadvantages of pie charts and bar charts were well known.

Question 2

A surprising number of candidates did not know the number of days in a year or whether 1881 was a leap year. (Neither omission was penalised.) Some assumed that (for example) a man recorded as aged 37 was actually aged exactly $37\frac{1}{2}$, which was not appropriate here. Many still do not appreciate that, to find the maximum possible difference between two rounded numbers, they need to work out the maximum possible value of the first minus the minimum possible value of the second, and to find the minimum possible difference they need the minimum of the first minus the maximum of the second. The midpoint of these two quantities is simply (max + min)/2.

Calculations were generally well done by those who remembered the formula for standard deviation. The negative numbers in the second calculation caused few problems, although many failed to note that the sum of squares remained the same. Some candidates obtained a negative value for the variance and failed to remark that this was impossible.

Question 3

Part (i) was generally well done except by candidates who failed to answer what was asked, and some who assumed that the times given were in hours and minutes. In part (ii), many candidates commented on the 6-hour limit before the penalty charge but few noted the unusual behaviour at 1, 3 and 4 hours. They need to consider the data more carefully. In part (iii), the median and quartiles were generally found satisfactorily, although many candidates ignored the information that each value had 5 as the units digit. (So, for example, the second largest value (stem 4, leaf 9) was to be assumed to be 495.)

In the last part, some candidates forgot to comment on the skewness.

Question 4

The unequal class intervals caused problems in this question. For the ogive, the scale on the horizontal axis must be uniform whatever size the class interval. The specification of the class intervals in the form '0.5 minutes but less than 1 minute' also caused difficulty for many in deciding where the upper class boundaries were. (They were at 0.5, 1, 2 minutes, etc.)

As indicated in previous reports, when calculating the position of the median for a grouped frequency distribution, 50% of the total frequency N is to be preferred, rather than 50% of $(N + 1)$, as is appropriate for ungrouped data. Candidates were not penalised, however, if they used $N + 1$ here.

The use of linear interpolation to estimate the median does not seem to be well understood, with many candidates trying to remember and use a formula that they did not really understand.

Most candidates made appropriate comments.

Question 5

This was generally the least well-answered question, as probability questions usually are! The best candidates made the point that using only 3 eggs will reduce but not eliminate the probability of food poisoning. The better candidates managed the calculations well. In the last part, a comment on whether independence was likely in this situation was needed, not just a statement of what the term 'independence' means.

Question 6

It was pleasing that this long question was well done by many candidates. The time chart was generally well drawn. Part (ii) was very poorly answered – most candidates seemed to assume that as it was four-quarterly data a four-quarterly moving average was needed. They need to be reminded to look at the chart they have just drawn to see what seasonal pattern, if any, emerges.

Most could calculate the appropriate moving average trend although a few did not centre and others did a repeat four-quarter moving average to centre. It was pleasing that very few candidates went on automatically to try to estimate seasonal variations, which were not asked for here. The calculation of the trend using regression was generally well done by those who attempted it. When plotting the regression line, candidates are reminded that it is a good idea to plot 3 well-spaced points or to plot 2 well-spaced points and check that the mean point lies on the line. The regression line should cover the complete time interval.

Question 7

It was probably easier, although not essential, to rank points scored from high to low and points conceded from low to high or vice versa, thus getting a positive coefficient.

Again, candidates had problems because they could not remember the formula for Spearman's rank correlation coefficient or remembered it incorrectly. In some cases, the calculated value for the coefficient lay outside the range $(-1, 1)$, but candidates did not comment that this must be wrong. Most commented on the value of the coefficient but only the best commented on the differences in ranking and on Ireland and France, in particular.

Question 8

This question was disappointingly answered. Many candidates seemed to be more familiar with indices for prices rather than costs, but the methods involved in the two cases are identical. It must be impressed upon candidates that when they state the value of an index number or a relative, they must also state the year for which they are calculating it and the year on which it is based, otherwise it is meaningless.

In part (iii), several chose to calculate an aggregate index rather than the weighted mean of cost relatives index requested.

Higher Certificate: Paper I – Statistical Theory

General

The aim of this paper is to test the ability of candidates to understand and interpret basic statistical theory and to apply and adapt it to simple practical situations.

The 2005 paper seems to have been perceived as being of about average difficulty. The range of performance was wider than usual this year. Whilst some candidates achieved impressively high standards, the work of a majority of candidates (those who failed or narrowly passed) showed many examples of serious mathematical weakness. Of particular concern are misuse of sigma and product notation and suffixes, inaccurate differentiation and integration, and confusion between discrete and continuous distributions.

There were few good answers to the parts of questions on conditional probability (Question 2 part (iv), Question 5 part (ii), Question 8 part (ii)). On the positive side, the highest average marks were obtained on Question 2 (Normal distribution), followed by Question 3 (Poisson-exponential relationship, distribution of sample minimum and sample maximum, albeit with rather few attempts) and Question 1 (combinatorial analysis, Poisson approximation to binomial).

Question 1

This question was popular, being attempted by 63% of candidates with an average score of 12.3/20. Parts (i), (ii) and (iii) were generally well done, but very few good answers were seen to part (iv). In this part, the number of lottery winners must be represented as a (single) binomial variate when at least one of the six winning numbers is greater than 31, and as a sum of two binomial variates otherwise. Each binomial distribution may be approximated by a Poisson distribution and, in the former case, the additive property of independent Poisson variates is then used.

Question 2

This question was easily the most popular, and was attempted by over 90% of candidates with an average score of 14.8/20. The first three parts were generally well done, although several candidates misread parameter values given in the question, or quoted Normal probabilities complementary to the ones required. There were a few good solutions to the final part, for which Bayes' theorem was required.

Question 3

Fewer than a quarter of candidates tried this question, and about half of these achieved good answers; the average score was 12.6/20. The question required derivation of the exponential distribution from the Poisson probability of zero, followed by riders on the distributions of the maximum and minimum of exponential random samples. Mathematical errors were common in the arguments given by lower-scoring candidates.

Question 4

This question was attempted by 80% of candidates, with only a few very good answers but disappointing results for many, giving an overall average score of 10.5/20. There were numerous poor answers for the cumulative distribution function in part (a), with many errors involving wrong signs, omission of limits of integration or the constant term, and leading to wrong graphs. The algebraic derivation of the maximum likelihood (ML) estimate was often garbled, and the derivation of the asymptotic Normal-theory confidence interval was seldom set out accurately. Several attempts used logs to base 10 instead of base e; other common errors were to centre the confidence interval on the mean of the given sample rather than on the ML estimate and, when calculating the endpoints, to apply a further divisor $\sqrt{n}$ to the standard error.

Question 5

This question was tried by fewer than a quarter of the candidates and yielded the lowest average score (7.9/20), there being just two excellent answers. Several attempts confused the initial set-up by assuming a sample of n binomial (n, p) observations, a muddle from which by no means all recovered. In part (ii), few candidates clearly understood the relationship between θ and p, and fewer

still correctly used the linear relationship to deduce the standard error of $\hat{\theta}$ from that of $\hat{p}$. However, in the final part most answers did recognise the potential of the anonymising device in the second survey to encourage more truthful answers.

Question 6

This question was attempted by about half the candidates, and achieved the second lowest average score (8.1/20), there being just three excellent solutions. In part (i), many graphs portrayed the geometric distribution as continuous. There were many errors, mainly failures of differentiation, in deriving the mean and variance from the probability generating function in part (ii). Several candidates were unable to give any convincing rationale for the probability of at least x failures in part (iii), or to interpret the given result as the 'lack of memory' property. Whilst some candidates successfully obtained $P(Z = z)$ in the final part, few were able to recognise Z as a geometric variate and thus to complete the solution.

Question 7

About two-thirds of candidates attempted this question; the average score was 10.2/20. The raw data were generally accurately plotted, although several answers made no reference to the fact that the trend was clearly nonlinear. With the aid of the summary data, the correlation was also usually accurately computed, but in part (ii) the linearising transformation was less often successfully handled.

Many candidates lost marks in part (iii), having not fully understood the logic of the transformation by which estimates of the parameters a and b could be recovered from the linear regression of Y on X.

Question 8

This popular question was attempted by 87% of candidates. However, there were only a handful of very good answers and the average score of 10.9/20 was disappointing. Part (i) was well done, with most candidates correctly deducing the discrete uniform marginal distributions of X and Y and their means and variances; however, a number rendered the variance as $E(X^2)$, failing to subtract the square of the mean. Performance in part (ii) was rather worse, with many answers failing to normalise the required conditional probabilities and very few correctly justifying the linear dependence of $E(Y|X = x)$ on x. Part (iii) was better, with many sensible attempts, albeit often with arithmetical errors. There were few good answers to the final part.

Higher Certificate: Paper II – Statistical Methods

General

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

Overall, the performance on the paper was quite good, with an average mark of a little over 60%. Many candidates could choose and apply a valid test and provide a brief interpretation. However, when a

more detailed explanation or justification of a methodology was required, or a report was required for a non-technical audience, the solutions provided were considerably more limited. Additionally, although a number of candidates achieved high marks for the parts of the questions requiring data or statistics to be represented graphically, the abilities of candidates to choose the required graph or, where no requirement was made, to select a form of diagram appropriate to the question was rather disappointing.

Many candidates, particularly those who appeared weaker in general, made errors in recalling the formulae relevant to the questions, even when they could identify the correct general approach to the analysis. However, candidates who could correctly recall the formulae were generally very good at performing calculations, although some did not provide enough detail on the method of calculation.

Candidates' weaknesses noted in last year's report were similar to those found amongst this year's scripts. There were several common errors in relation to tests of hypotheses:

- using a test for independent samples instead of one for paired data, when the samples were clearly related;

- for tests where a 5% significance level and both tails of the relevant sampling distribution (under the null hypothesis) were needed, using the 5% points rather than the 2.5% points.

On a related matter, it will also help candidates if they always bear in mind that the purpose of a hypothesis test is to assess the strength of evidence *against* the relevant null hypothesis (H_0). That hypothesis should be rejected when the evidence is sufficiently strong. But, if the evidence is not strong enough to justify rejecting H_0, that does not of course prove that H_0 is true. For this reason, it is better to report 'we do not reject H_0' rather than 'we accept H_0'.

Question 1

This question was not particularly well answered. In particular, the discursive parts of the question were not well addressed. Relatively few candidates showed a good understanding of why the Poisson distribution might be appropriate to model the number of goals scored by the home teams and no candidate suggested the negative binomial distribution if the variance had been considerably larger than the mean. Worryingly, several candidates suggested continuous distributions for the latter case. It is important for candidates to take account of the nature of any distribution suggested for modelling a particular situation; both the type of random variable (continuous, discrete, categorical) and any range restrictions (e.g. non-negative values) should be considered.

Most candidates correctly calculated the sample mean and variance. Additionally, most candidates calculated the expected frequencies correctly, although candidates should be reminded to take care with rounding errors. Special care is needed in cases such as this where a small value (such as the Poisson mean or expected probability) is multiplied by a large value (the total frequency) to give the expected value, which is then subtracted from a similar value (the observed frequency). Many candidates, but not all, correctly noted that estimating the Poisson mean results in the loss of an additional degree of freedom. Interpretation of the results could have been better; the distinction between accepting the Poisson hypothesis and not rejecting the hypothesis could have been better made by many candidates. A good way of expressing this conclusion would be to say that 'the Poisson hypothesis appears consistent with the data'.

Question 2

This was the most popular question, and the standard of answers was consistently good. Only one candidate failed to achieve at least 10 marks out of 20 on this question.

Part (i) was answered well, but candidates are reminded to justify how they find the median and, in particular, the quartiles. Several candidates incorrectly defined an outlier in the right tail as being above $1.5(Q_3 - Q_1)$, or even above $Q_2 + 1.5(Q_3 - Q_1)$, rather than above $Q_3 + 1.5(Q_3 - Q_1)$, which was specified in the question as the convention to be used. Several candidates correctly drew a boxplot, but did not note that the distribution was positively skewed, which affected their ability to answer part (ii) satisfactorily. Several candidates unnecessarily wasted time performing a t test for part (ii). Candidates should ensure that they read the question carefully: no test was required.

Part (iii) was answered quite well, although few candidates noted that the method of sampling by the audit department was not stated in the question. Testing was based on the assumption that random sampling was used; since this may well have not been the case, the result may not be reliable.

Question 3

This question was quite well answered. Some candidates chose to perform a χ^2 test for a 2×2 table (some, but not all, using Yates' correction); others chose to test the equality of two (binomial) parameters, using a Normal approximation. These approaches are all acceptable, although the χ^2 test with Yates' correction is often preferred.

Part (ii) was well answered, although few candidates presented the results as a percentage and, likewise, few made it sufficiently clear (in words) which way round they calculated the difference between the proportions (that is, *male − female* or *female − male*). Whilst a number of candidates recognised that the sample sizes limited the validity of the Normal approximation, many did not realise that it was the sample sizes of men and women rather than the size of the total sample which was an issue. With samples of size 20 to 30, as here, the approximation will tend to be moderately good, but will be less so when (as with the males here) the sample size is small and the asymmetry of the underlying distribution more pronounced. Here the distribution was binomial, and the proportion was relatively far from 0.5, resulting in a very skew distribution.

Question 4

This question was generally well answered, particularly part (ii).

In part (i), a number of candidates incorrectly used a χ^2 test for a 2×2 contingency table; McNemar's test was required here.

Whilst most candidates correctly computed the standard error of the proportions in part (ii), a number of candidates appeared to be confused between the standard error of an estimate of a proportion ($\sqrt{p(1-p)/n}$, in conventional notation) and the standard deviation $\sqrt{np(1-p)}$ of a binomial random variable X. Most candidates who confused these did not state clearly what the formula was intended to represent; the first is the standard error of $\hat{p}$ and the second is the standard deviation of X. Since $\hat{p} = X/n$, it is clear that $\mathrm{Var}(\hat{p}) = \frac{1}{n^2}\mathrm{Var}(X)$, but candidates who confused these sometimes divided the expression a second time. Clear descriptions and labelling of calculations can reduce the risk of confusion.

It is noted that, whilst this was, indeed, a relatively short question, many candidates still included insufficient details of their calculations.

Question 5

Most candidates obtained good marks on this question, largely because part (ii) was answered well.

Almost all candidates correctly calculated the sample means and variances from the summary statistics provided. Many candidates also calculated the value of the test statistic accurately, and correctly referred this to the 2-sided percentage point of t_{19}, choosing not to reject the null hypothesis. We have commented earlier about how such a result should be interpreted; in the present application the most appropriate conclusion is that there is no evidence to suggest that the mean house prices in the two towns differ.

In part (i), most candidates noted that the assumptions included those of Normality and equality of variances (homoscedasticity). However, it should be noted that these assumptions relate to the underlying populations (or distributions) rather than to the samples. Related to this is the assumption that the data were (independent) random samples from the underlying populations: few candidates stated this assumption. Additionally, under the null hypothesis, the means of the two populations are equal; it is of course this hypothesis which is tested in the t test for two independent samples.

Part (iii) was not well answered, although some candidates, typically those who scored quite highly throughout the paper, correctly identified the main issue. This was that the samples could not reasonably be assumed to be randomly selected from the populations of two-bedroomed terraced houses in the two towns. Other candidates identified that the validity of the test could be affected by one or more of the assumptions not holding whereas others referred to other issues unrelated to this. It is important that candidates recognise:

- the assumptions underlying the derivation of any inferential statistical method; and
- that the validity of any test is affected by whether these assumptions hold, together with the robustness of the test to the likely type and degree of any departure from these assumptions.

Question 6

This question was quite well answered.

In part (i), candidates generally noted that the Mann-Whitney (or Wilcoxon) test might be preferred in cases where it was unreasonable to assume Normality. Not all candidates noted, however, that the t test would still be acceptable if the sample size were large, and that this test might be preferred to the Mann-Whitney test in such cases if inferences regarding the comparability of the underlying population *means* were of interest. Additionally, few candidates noted that the Mann-Whitney test would be preferred in cases when it was unreasonable to assume that the data were on an interval scale but where it would be reasonable to assume the scale to be ordinal (and therefore that ranking the data would be a valid action).

Many, but not all, candidates drew a dotplot satisfactorily. Most candidates noted that one or both of the distributions appeared skewed and concluded that a t test might be invalid. However, disappointingly few noted that the sample sizes were too small to appeal to the Central Limit Theorem for the validity of the t test. The Mann-Whitney test was applied satisfactorily in most cases.

Question 7

This question was moderately well answered. Part (i) was answered quite well by many candidates, although several included a term for 'blocks' which, as this would be a second factor, would lead to a two-way analysis of variance. Most candidates performed the analysis of variance correctly.

Relatively few candidates correctly performed the post-hoc testing required. It is important to follow up a finding of significance from an analysis of variance by determining where the differences between the population means appear to lie, and it was explicitly required in this question.

The final part of the question required a brief report for the researcher. It is a vital skill of an applied statistician to be able to write a report, giving valid conclusions and recommendations expressed in terms appropriate to the researcher requesting support. In this case, most candidates did not produce a report which would have been of use to the psychologist.

Question 8

Only a few candidates attempted this question. The standard of solutions was lower than for the other questions, although the candidates choosing to answer this question tended to score relatively poorly elsewhere on the paper.

For part (i)(a), candidates should ideally have used a time-series chart, or some other chart which illustrates trends, preferably using the same scale for males and females. For part (ii)(b), various types of diagram were used by candidates to provide illustration, many of which failed to provide a satisfactory representation of the data for the required purpose. Some candidates used pie charts, and one used a time-series chart of the absolute differences; note that the former enables the reader to assess only relative differences whereas the latter enables one to assess only absolute differences. The use of a clustered bar chart or 'time-series' chart was a satisfactory mode of representation. Some candidates also did not use the 1998 data; it is vital that candidates ensure that they read and address the question posed.

The interpretation of the graphs was generally barely satisfactory. This was due, in part, to the limitations of the diagrams. Candidates tended to focus attention on the absolute differences between males and females. However, particularly when comparing age groups, relative differences were also of potential interest. (For example, the average consumption of men was two to three times as much as that of women, and this ratio was higher for the older age groups.)

Higher Certificate: Paper III – Statistical Applications and Practice

General

The aim of the syllabus of Statistical Applications and Practice is to develop skills in data analysis, using the theoretical concepts developed in the syllabuses for the Ordinary Certificate and the other Higher Certificate papers. A principal objective is the analysis of data and the effective and comprehensive communication of the results. The questions on this paper require candidates to select and carry out appropriate statistical procedures, computational and graphical, and to report the findings and conclusions clearly.

Questions 1, 2 and 3 were the most popular, and were attempted by the vast majority of candidates.

Question 7, attempted by just under half the candidates, in general attracted the best performance. Question 8 was least popular; this and Question 4 were answered least well.

Question 1

Most candidates appreciated the concept of interaction but most did not understand how to assign degrees of freedom to the different treatment effects.

Most candidates did not know how to calculate the variance of a treatment effect.

Most candidates had some appreciation of the fundamentals of the differences in using an experiment with 16 rather than 4 rabbits, but many failed to mention that using fewer rabbits several times each leads to more precise results (that is, the variances of the estimators are smaller).

Question 2

Most candidates noted the assumption of Normality, but some did not mention the assumption that the population variances are equal, which is the basis for the use of a pooled estimate of variance.

A two-tailed test was frequently used instead of the indicated one-tailed test.

In part (i)(b) many candidates correctly identified that the Normal distribution and the previous estimate of the variance of the difference of means was appropriate. In this part of the question, candidates needed to examine the case in which the difference between the population means was 0.5; but many candidates set the difference between *sample* means at 0.5 or used the sample means from part (i)(a) in some way. Few obtained the correct value for n or realised that the total number of cars needed was $2n$.

Question 3

The practical interpretation of the linear regression model asked for in part (ii) was often poorly expressed. Many candidates noted the negative value for α, but did not comment on implications for the model or the practical relationship between PS and SM.

In part (iv), the variance of the estimated mean of Y given x (that is, of $\hat{\alpha} + \hat{\beta}x$) was frequently erroneously given as σ^2/n. Only a few candidates seemed to appreciate that it was a confidence interval for a mean which was required; naturally, the individual observations at $x = 80$ would cover a wider range. Also, critical values from $N(0,1)$ were frequently used rather than from t_{36}.

Question 4

Most answers failed to identify the average response and the trend as separate components of a time series. Few candidates were able to discuss weighted versus unweighted averages or suggest weight variations. Many candidates misunderstood what they were looking for in the irregular component plot.

Question 5

Many candidates scored well in parts (i) and (ii) but did not interpret part (iii) correctly. The concept of 100% inspection was misunderstood. Credit was given to those who assumed that in the first scheme the original sample remained identified and left 980 as the residual population, as long as an adjustment was then made to p to allow for the fact that fewer defectives would remain to be detected. The same comment applies to the second scheme.

Question 6

Many candidates were able to make sensible comments about the sampling situations. Many answers though were too long - good points could have been made quite briefly. In part (iv) a confidence interval for the difference in proportions was often given. A one-tailed significance test is a better answer.

Question 7

A number of candidates ignored the instruction to write down the likelihood function in terms of the mean rather than in terms of the exponential rate parameter. Whilst it is still possible to derive the maximum likelihood estimator required, it is better to follow the instructions given in the question.

There were often significant numerical errors, either in the calculation of components making up the χ^2 statistic, the degrees of freedom or use of the statistical tables.

Question 8

This was the least popular question and attempts were generally poor. Some basic and key features of the problem were often overlooked (for example, that the product was meant to be 'square') and some very practical issues were missed.

Graduate Diploma: Statistical Theory And Methods I

General

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

The overall standard of attempts at this year's paper was good. However, three candidates found only four questions they could attempt.

There were reasonable numbers of attempts at all the questions, but those on conditional probability (Question 1), order statistics (Question 6) and Markov chains (Question 8) were somewhat less popular than the others.

In general, candidates seemed more comfortable reproducing standard proofs than solving problems.

Question 1

This examined conditional probability (Law of Total Probability and Bayes' Theorem, extended to the use of iterated expectation and variance). About one-third of candidates attempted this question, but with very mixed success. Most were able to state the standard results required in part (i), but encountered real difficulty when tackling part (ii)(a). This required an application of the Law of Total Probability, but several candidates were unable to express the problem in the form required to apply the result.

Question 2

A majority of candidates tried this question, which tested their knowledge of how to add independent binomial random variables. Several candidates achieved virtually full marks, whilst others gained almost no marks. Many of them struggled with part (ii) and were unable to write the conditional

probability $P(X = x|X + Y = z)$ in a form they could handle. Those who reached part (iii) generally completed it successfully.

Question 3

Almost every candidate attempted this question, which was on jointly continuous random variables. Their answers were generally of a high standard. Several candidates scored all the marks for part (i) but were unable even to start part (ii); they could not identify the region corresponding to the event $Y - X > z$. Those who struggle with this type of question would find it extremely helpful to draw a rough sketch of the joint range space.

Question 4

This examined transformations of jointly continuous random variables. About two-thirds of candidates attempted this question and they did very well. Candidates should note that $f_{XY}(x, y)$ factorising into a function of x times a function of y does not guarantee that X and Y are independent; the joint range space must also be the Cartesian product of the marginal range spaces. (The range for X must not depend on the value of Y, and vice versa.) In general, in this and some other questions, candidates tended to ignore range spaces altogether.

Question 5

About two-thirds of candidates attempted the question, which examined the use of moment generating functions. Most of those who tried this question did reasonably well. Several candidates had trouble with the integration in part (i).

Question 6

This was a standard exercise on order statistics. About half the candidates attempted this question, and they obtained very good marks for it. As in Question 4, many candidates avoided writing down the range spaces of the variables they constructed; this caused problems to several of them.

Question 7

About two-thirds of candidates attempted this question, on simulation. The general standard of answer was very high. A few candidates had a problem obtaining negative random variates in part (i)(a) and, in part (i)(c), several candidates did not realise that a χ^2 variate could be obtained by squaring a standard Normal variate.

Question 8

This question examined Markov chains. Only one-third of the candidates attempted this question, and they obtained a wide range of marks. Several candidates did not appear to be able to invert a 2×2 matrix or multiply 2×2 matrices together.

Graduate Diploma: Statistical Theory And Methods II

General

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

All but two candidates answered exactly five questions. Of the remaining two, one attempted just one

question and the other answered six. Answering more questions than the rubric allows is usually not a good strategy: it is better to try to answer five questions well.

Almost all candidates showed at least some understanding of the material and a majority showed a good understanding. The overall level of preparedness of the candidates appeared to be considerably greater than in recent years. Many candidates gathered marks efficiently, showing good examination technique. There were some strong performances.

In the main, candidates appeared to be strongest at likelihood-based material (both frequentist and Bayesian). The work on non-parametric methods was less good overall. Sufficiency is an area where candidates' knowledge tended to lack conviction. Overall, candidates did better at manipulation than interpretation.

Question 1

All candidates attempted this question. All candidates had the right idea for part (i) but a few did the integration incorrectly. Answers to part (ii) were generally good. There were many good answers to part (iii) but there were occasional difficulties with the form of the Cramér-Rao lower bound. In part (iv) the MLE was usually obtained correctly and there were some reasonable brief explanations. A few candidates wasted time trying to evaluate the expectation explicitly.

Question 2

Over 80% of candidates attempted this question. Part (i) was usually done successfully. In part (ii) most (but disappointingly not all!) calculated the mean of four numbers correctly but few spotted the fact that the estimate is useless because it is less than the largest observation. There were some reasonable attempts at part (iii).

No candidate got to the end of part (iv). Most did the necessary moment calculations and knew what MSE is. However, there was a tendency to introduce the special case too early in the calculations; leaving the formula in general form until the final step makes the algebra so much simpler.

Question 3

Just under half the candidates attempted this question. Parts (i) and (ii) gave the opportunity to earn 9 marks very easily simply for stating standard results for sufficiency. Unfortunately, very few candidates knew this bookwork. In part (iii) most obtained the likelihood correctly but application of the factorisation theorem was weak. Part (iv) was mostly done well.

Question 4

About a third of candidates attempted this question. Answers to part (i) were mostly too vague to get high marks. There were few serious attempts at parts (ii) and (iii) but these tended to be good. There was only one decent attempt at part (iv).

Question 5

Over 80% of candidates attempted this question. Parts (i) and (ii) were generally done well. There were some good answers to part (iii) but a common error was not to realise that bias and MSE involve taking expectations over the observations; that is, they are repeated sampling properties of the Bayes estimator.

Question 6

Around three-quarters of candidates attempted this question. Parts (i) and (ii) were well done, though a few candidates confused variance and standard deviation. Part (iii) was disappointing, especially most candidates' inability to plot the pdf of an exponential distribution. Again variance and standard deviation were sometimes treated as the same. Part (iv) was mostly well done.

Question 7

Around a third of candidates attempted this question. The attempts were generally serious, and there were some good points made. However, the style of writing tended to be rather rambling, and as a result some important points were missed altogether. Other points tended to be poorly explained; for example, statements about expected sample size in part (a) and about power in part (b).

Question 8

Around a third of candidates attempted this question. All got the result for the mean in part (i) but no candidates obtained the variance result correctly. There were few serious attempts at parts (ii) or (iii) but those that got this far at least made some progress.

Graduate Diploma: Applied Statistics I

General

This paper tests an understanding of the basic theory underpinning a range of statistical methods, notably general linear modelling, time series, multivariate methods and generalised linear models. It also tests the candidates' knowledge of how to apply these methods in realistic applications and how to interpret computer output in the context of these applications.

There were two main weaknesses which resulted in loss of marks. Firstly, candidates did not always answer the question being asked, but produced bookwork which had some relation to the question but failed to answer the specific question in the context in which it was asked. The second, related, problem was that although candidates could on the whole replicate theory, they were much less able to relate the theory to practical applications.

Question 1

This tested basic theory about MA(1) and AR(1) time series. The theory was reasonably well reproduced; parts (iv) and (v), which were more practical, were not so well done.

Question 2

This was a question about general linear models, and the use of factors in these models. Although candidates were able to distinguish between a factor and a continuous variable they showed little understanding of how a factor can be coded in a linear model, and many could not produce the design matrix asked for.

Part (b) of the question was about correlated predictor variables. Answers were rather sketchy and lacking in detail, although most candidates showed a basic grasp of what the question was about.

Question 3

The first part of this question was standard theory, which was reasonably well presented, except that

several candidates omitted parts of the answers for no apparent reason. It is important to read the question carefully and to provide an answer to each part.

Candidates were not very good at interpreting the residual plots, which was the more practical part of the question. Examples like this are given in text books, and candidates should be familiar with how to interpret them.

Question 4

This was a question about model selection in multiple regression. Candidates were quite good at describing methods, but were less good at relating methods to the practical scenario posed. Answers to part (ii) tended to be rather brief and did not show a full understanding of the statistics presented. Very few candidates appreciated the importance of a wide-ranging approach, taking into account the nature of the variables as well as the goodness of fit of a model.

Question 5

This question tested a basic understanding of a generalised linear model for a binomial distribution. Very few candidates attempted the question and the answers were poor. It is important to know how to select a model based on the scaled deviance. It is also important to understand the role of the linear predictor in such a model.

Question 6

Answers to this question were poor. Even the first part, which asked for basic definitions, was answered badly. Few candidates could describe the proposed distance measure in mathematical terms, and so check whether it was a valid distance measure. The latter part of the question was about the practicalities of cluster analysis; this technique can be very subjective, and few answers showed any appreciation of this.

Question 7

As in previous years when there has been a question on MANOVA, few candidates attempted this question, and those who did try it showed little understanding of the theory underlying the method. MANOVA is clearly on the syllabus!

Question 8

As usual, this question was popular. However, also as usual, not all candidates correctly specified the model. Even those who correctly identified the random and fixed effects did not always interpret the results correctly. Although answers to questions of this sort may require a lot of arithmetic, marks will still be lost if the interpretation is not correct, since this is a key feature of an applied statistician's role.

Graduate Diploma: Applied Statistics II

General

The Applied Statistics II syllabus covers the fundamental concepts of designed experiments and sample surveys, and their application to data.

All candidates followed the rubric, and most attempted exactly 5 questions. As in previous years,

some candidates lost marks by not answering the question asked or omitting sections of the question entirely. Some candidates presented bookwork that seemed relevant to the general area even though it was not related to the specific question asked: this is not a good strategy, since the examiner's marking scheme cannot award marks for this material. Some candidates wrote just a few lines on each question, and were clearly not ready to sit this paper.

The overall standard of answers was, unfortunately, generally quite poor. Two candidates performed exceptionally well, but the average mark per question was generally below 10. It was pleasing, though, that the demography question was answered well, with an average mark of 13.

As was the case last year, candidates seemed more comfortable when reproducing standard book-work than when dealing with the practical aspects of surveys and experimental design. Fundamental concepts were often poorly understood. Topics such as response surface methodology were again avoided, and were only attempted by the more able candidates. Candidates should be encouraged to gain a deeper understanding of all the topics on the syllabus, and of their application, since this paper is unlikely to be passed on standard bookwork alone.

Question 1 (14 attempts)

The concept of missing data, the effect on the design structure, and the assumptions which must be made for the analysis to be valid, were often poorly understood. Candidates gave confused or wrong answers, or simply omitted parts (i) and (iii).

Parts (ii) and (iv) were reasonably done. Some candidates made mistakes in calculating approximate standard errors for the difference between diet means in part (iv), either incorrectly stating the degrees of freedom for residual mean square, or using $\sqrt{\frac{2}{6}}s$, where $s = \sqrt{3.7111}$, as the standard error for all treatment comparisons. Significance tests were generally performed satisfactorily, and were interpreted correctly, though candidates should pay more attention to stating the conclusions.

Hardly anybody commented on the last part of part (iv). Candidates should have realised that the treatments were quantitative, and further analysis using polynomial contrasts might have been useful.

Question 2 (10 attempts)

In part (i), most candidates could explain the concept of confounding in 2^k factorial designs, and could construct a simple confounded design.

Part (ii) was not done well. Almost all candidates overlooked the fact that there were two replicates of the experiment. Few candidates were able to identify the confounded contrasts in part (ii)(a), or correctly construct the analysis of variance table in part (ii)(c). Often, candidates listed blocks and the interactions confounded with blocks as if they were separate sources of variation.

In part (ii)(d), candidates were required to calculate the standard error of the factorial effects. However, some gave the residual mean square as the answer, and then wasted time by unnecessarily calculating the analysis of variance (which was not required).

Part (ii)(c) of the question included output showing the mean yields for each two-factor combination of a 2^4 factorial design. Most candidates seemed unfamiliar with seeing the data presented in this format. Even though the best way to calculate a main effect was explained in the question, many candidates calculated the main effect of A in a more complicated way: they averaged results from all

the two-way tables involving A, although only one of these tables was needed. Most candidates were unable to verify the effect for the BC interaction.

Question 3 (16 attempts)

It is disappointing that such fundamental principles as replication and randomisation in experimental design are so poorly understood. Candidates could explain these principles, but not necessarily their importance. Few candidates appreciated the importance of including a 'control' arm in a study aiming to evaluate the effect of a treatment.

Part (ii) was done well. Most candidates were able to perform a standard analysis of variance for a 3×2 factorial design. Some candidates lost marks by omitting the sum of squares (SS) for interaction. Almost all candidates presented means for all combinations of the levels of the two factors, even though the interaction SS was not significant. When an interaction is not significant, it is valid to examine just the main effects, and express results in terms of these alone.

Few candidates questioned the assumption of constant variance in part (b), in spite of the somewhat skewed appearance of the data recorded. Hardly anybody commented on the presence of zeros in these data; this is a very good reason for not applying a simple logarithmic transformation.

Question 4 (5 attempts)

It was disappointing to see so few candidates attempt this question, which was a relatively straight-forward application of a fractional factorial design for fitting a first order response surface. Some answers were quite good.

Question 5 (7 attempts)

This question examined candidates' ability on a range of topics including random and non-random sampling methods, practical aspects of sampling, sample size and properties of estimators.

Often, answers to part (i) were too short or incomplete. There were some good examples of volunteer sampling. Most candidates gave satisfactory reasons in part (ii) for preferring systematic sampling to simple random sampling in this case. However, many candidates clearly had little or no idea of how to calculate a sample size for a simple random sample, and of the circumstances under which the simple random sampling formula would be appropriate with systematic sampling. Those who attempted part (iii) gave satisfactory answers for the proportion of schools that have a playing field, but failed to recognise that students within schools represented a cluster sample.

Question 6 (11 attempts)

Parts (i) and (ii) were bookwork on stratified sampling. Most answers were satisfactory.

Part (iii) was not done well. Candidates were required to compare two designs for selecting a sample of households, subject to a fixed budget for the survey. Many candidates were very confused over what formulae to use to calculate the variance of the estimated population mean. Also, the population size N was not specified, and candidates failed to note that the variance terms should be expressed as a function of N.

Design 2 used stratified random sampling. The minimum variance for fixed cost is obtained by substituting n_h from part (ii) into the general formula for $\mathrm{Var}(\bar{y}_{st})$. Most candidates gave confused or

wrong answers or did not complete the question. Design 1 used simple random sampling. Many candidates had difficulties estimating the population variance S^2. Several candidates incorrectly used a formula such as $\hat{S}^2 = \sum S_h^2 / N_h$, which overlooks the fact that the stratum means differ.

Question 7 (7 attempts)

Part (i) was standard bookwork on simple random sampling and its application to selecting a sample from an electoral register. There were some quite pleasing answers. Part (ii) was not done well. Those candidates who attempted parts (b) and (c) clearly did not know how to calculate a regression estimator of the population proportion, or understand in what way the method uses auxiliary information to adjust the outcome of interest.

Question 8 (12 attempts)

This question on population pyramids and age-specific death rates was generally done well, and attracted some of the highest marks on this paper. Marks were lost on part (c), because the similarities and differences in mortality rates were not always clearly stated.

Graduate Diploma: Options Paper

The numbers of candidates for most components of this paper are small, and it is not possible to give detailed reports without identifying individual answers.