ROYAL STATISTICAL SOCIETY EXAMINATIONS, 2001

REPORTS OF EXAMINERS

Ordinary Certificate Paper I

On the whole Paper I was well done this year with an overall average of 60.7 and a 75% pass-rate. The main weaknesses of the candidates were the numerical sections.Many could not work out the required percentages in Question 1 and many had trouble finding the required sample sizes in Question 5.

Candidates did not always follow the instructions on the front of the answer book. In particular, the following caused difficulties this year.

1. Begin each answer on a new page.

2. Write on both sides of the paper (to avoid having unnecessary loose sheets of paper).

5. Be sure to fill in numbers of questions attempted at the bottom of the first answer book but not to fill in the side panels.

8. Make sure handwriting is legible.

Question 1

Most candidates attempted this question and the average mark was more than half the total allotted mark. The main problem was that some candidates did not realise that the sample was only being chosen at age 16 and then followed over the subsequent time periods.

Candidates were not able to carry out the calculations required in part (iii).

Question 2

This was a popular question and many candidates scored well and drafted logical and easy to use forms. There were some good individual touches with suggestions for suitable prizes and disqualification of photocopied forms. Several forms included a data protection statement. Several forms did not indicate where to send the entry. Many assumed that only one paper was bought on Sunday. Some good answers suggested that requesting a forename rather than a Christian name might be diplomatic. Several made the form longer than necessary by asking how many times a week the paper was bought and then asking which days it was bought.

Question 3

This was a popular question where it was easy to score some marks but difficult to get full marks. Many candidates answered the question in general terms rather than relating it to the specific electoral roll context. In this instance, a sampling frame already exists and it is already stratified by ward so the costs associated with stratification are minimal. Worries that some strata would be too small were unfounded. Some candidates gave a general description of the methods rather than answering the question set. Many assumed that in the quota sampling, equal numbers would be sampled from each ward. Even though the question specified that quotas were based on wards, many seemed to assume that quotas based on age group, gender (i.e. sex) and income group were automatically required as well. A few confused stratified sampling with multi-stage sampling and assumed only a sample of wards were to be surveyed.

The consequences of the electoral roll being out-of-date were generally well appreciated and candidates seemed to have been well prepared on this point.

Question 4

Almost all candidates attempted this question and the average mark was high. Almost all knew the additional number needed in part (i) and gave an appropriate extra question in part (iii). Many could list three problems associated with the sampling method but surprisingly few picked up on the fact that the percentage yes vote was announced part way through the time period producing a very obvious source of bias. Many did however point out that dog owners might be out walking their dogs at the vital time!

Question 5

Again all but a handful of candidates attempted the question and the average mark was quite high. Many defined the reciprocal of the sampling fraction. Most candidates could work out the required numbers in the sample using a uniform sampling fraction but many were not able to complete part (iii). A large number used number in sample proportional to standard deviation rather than sampling fraction proportional to standard deviation.

Question 6

This was a popular question but many candidates failed to appreciate the practical difficulties of observation. It is unreasonable to expect to observe more than one customer at a time. Many students got sidetracked into trying to get samples that were of equal size and age/gender composition on the two days. Although sex can be easily observed, age cannot. A few candidates suggested interviewing customers rather than observing them. Specific details of how times and value of purchases were to be observed were lacking in many answers. There were some excellent answers that described the choice of customer appropriately and suggested use of stopwatches and till rolls. Some noted that the use of CCTV, if installed and able to cover the whole shop, might be valuable here. On the whole the answers were too long for the number of marks allocated to this question.

Question 7

This question was attempted by about 90% of candidates and the average mark was high. It was a bookwork question and candidates had mainly learned the bookwork well. There was some confusion in part (iv) between ways of trying to minimise non-response and ways of adjusting for non-response bias.

Question 8

A considerable number of candidates did not attempt this question. Many answers did not specify field types and widths. Others consisted of a form, which was not asked for. Some answers indicated that the candidates were familiar with a relational database such as Access and they tended to answer the question well. It is usually good to split up the various elements of a name and address and input them separately as this allows for greater flexibility in searching and using the database.

Ordinary Certificate Paper II

As in previous years, the paper was designed to test candidates' understanding and interpretation of statistics as well as checking on their technical ability to perform standard calculations and to draw appropriate graphs. The majority of candidates attempted all questions although a substantial number

attempted only the more routine parts of questions and omitted the interpretation.

It would be helpful if candidates could remember to study the instructions on the front of the answer book, concentrating in particular on the following:

1. Begin each question on a new page.
3. Write on both sides of the paper.
4. Attach graph paper opposite the answer to which it relates.
5. Fill in the numbers of questions in order attempted in bottom panel of first booklet (not side panel).
6. Legibility and clarity of expression (particularly when using technical terms)

It is also important, and very much in candidates' interests, to make absolutely sure that all extra sheets are firmly attached.

The overall standard was disappointing this year. This perhaps reflects the fact that there was less routine computation on the paper this year with more emphasis on interpretation and understanding. The charts and graphs requested also required more understanding and Question 3 (cumulative percentage frequency polygons) was poorly done in general, although the scatter diagram in Question 6 was better.

Probability is still difficult for many – encouragement to use tree diagrams and Venn diagrams, where possible, would seem to help understanding.

Many candidates were short of time, but often this was self-inflicted through lack of understanding/knowledge. For example, in Question 4, there was no need to convert to degrees Fahrenheit and, in question 7, some reworked the trend values even though these were given in the question.

Question 1

A substantial number of candidates did not attempt this question and the average mark was very low. Some candidates did manage the linear interpolation (possibly by drawing a graph) but very few could express the resulting answer as a weighted combination of the given p values.

In part (ii) there was still much confusion about writing down the minimum and maximum possible values for a value which is given correct to a number of decimal places. Some candidates insisted on writing the maximum value as 0.76424999999 and others appeared to think that either the maximum or minimum value must end in a 4 rather than allowing them both to end in 5.

Not surprisingly, very few managed the last part of the question.

Question 2

Most candidates attempted this question and all but a handful were familiar with a stem-and-leaf format but many students failed to obtain full marks because of carelessness in the execution of their diagrams.

There are several different formats that are acceptable for the diagram as shown in the suggested answers but two points are crucial – the leaves must be vertically aligned and the units of both stem and leaf must be stated clearly. The units of the stems must be constant – some candidates used a stem of 10 for values below 100 seconds and a stem of 100 for 100 seconds and above. If two digits

are used for the leaves, then all leaves must be displayed with two digits including those under 10 (00,01,..., 09). Some candidates attempted to use a stem of 10 throughout but they did not remember to include those stems with no leaves. This choice of stem leads to an unsatisfactorily spread-out diagram.

There was no need to change the data to minutes and seconds for the diagram, although at the comments stage the use of minutes often helped understanding.

Most candidates identified the modal stem but they were careless in the wording of their comments about it. It is not true to say that *most* of the calls were between 200 and 299 seconds or that the *majority* of calls were between 200 and 299 seconds. A handful of candidates failed to comment on this specific stem-and-leaf diagram and instead wrote in general terms of what these diagrams are meant to achieve. The best answers related the diagram to the call centre situation, identifying three groups of calls: those under 3 minutes (perhaps just an enquiry); those between 3.5 and 5 minutes (perhaps a standard order or repeat customer) and those over 6 minutes (perhaps more complex orders or new customers) and suggesting that further work was necessary to characterise these groups.

Question 3

Several candidates made no attempt at this question. This was surprising in view of the number of marks (16) it attracted. Some, who were short of time or unsure of how to draw the graph, attempted the final part only and were able to gain a few marks for suitable comments.

The usual comments about all graphs apply here: that they should have a heading, clearly labelled axes and units, ruled axes and a key or legend. As the graphs for Company A and Company B intersected, the two lines need to be distinguished by weight of line, colour or type of marker. Almost one third of attempts had an incorrect scale or no scale at all on the horizontal axis, presumably because candidates were used to drawing cumulative frequency polygons for equal-interval distributions. The scale has to be an arithmetic scale and the 'less than' cumulative frequencies are plotted at the upper boundary of the class interval not the midpoint. It was appropriate to add the point at the lower end of the distribution to indicate that 0% of leavers left after 0 months service. It was not appropriate, however to include a point to indicate the maximum months of service 100% of the leavers had served – this information was not given in the question.

Most who drew the graph knew how to obtain the median and quartiles from it but did not always read their scales accurately. Some worked out the quartile deviation rather than the interquartile range.

A few candidates drew inappropriate graphs – frequency polygons, stacked bar charts and cumulative 'histograms'. Some did not convert the frequencies to percentages.

Question 4

Most candidates attempted this question and the average mark was slightly more than half the total mark. Most could give an advantage and a disadvantage of using the standard deviation as a measure of spread.

In part (i) a surprising number of candidates calculated the standard deviation using divisor n rather than $(n-1)$ even though the data given were clearly a sample. They were not penalised for this.

In part (ii) most gave themselves much extra work by converting the data to degrees Fahrenheit and calculating the mean and standard deviation of the new data instead of using the appropriate formulae for finding the mean and standard deviation of one variable which is linearly related to another. Those who could use the formula for the mean correctly mostly used exactly the same formula for the standard deviation forgetting that the addition of 32 is irrelevant as far as the standard deviation is concerned.

Part (iii) seemed to puzzle all but the very best candidates; some of the others could obtain a mean of zero but could not cope with the unit standard deviation. Candidates should remember that, here, the range is the largest value minus the smallest value.

Question 5

This question was attempted by about 90% of the candidates and the average mark was almost exactly half the total available. Candidates still seem to have troubles with elementary probability. The better candidates drew a tree diagram (rather than a Venn diagram which would have worked just as well) to help with their calculations.

The idea of conditional probability seemed not well understood both in (ii) (b) and also in part (iii). Few candidates realised that, in calculating the probability of an accident, the tree diagram needed an extra layer.

Question 6

This question was again attempted by about 90% of the candidates, however better marks were obtained by those who did attempt it and several candidates scored very highly on this question.

Some candidates had trouble writing down pairs of rankings to give the maximum and minimum values of the coefficient. Almost all the candidates could draw the scatter diagram but some were careless in plotting and some did not indicate the outlier. On the whole most candidates could calculate Spearman's rank correlation coefficient for the complete data set though some did attempt to use the actual values to find the required differences. When the outlier was excluded several candidates did not recalculate the rankings but just removed the difference for the outlier from their calculations.

Question 7

Almost all candidates attempted this question.

Part (i) was very poorly done. Very few could explain the term 'seasonal component' without tautology. Many stated when it was appropriate to use an additive or a multiplicative model but did not say what the models are.

Part (ii) was generally well done although some candidates worked out the trend values even though they were given in the question. Most laid out the calculations appropriately and understood the need to adjust the averages, although several adjusted in the wrong direction.

In part (iii), a few applied the seasonal variations in the wrong direction and some did not use the estimates of trend given in the question. In both parts (ii) and (iii), very few candidates gave the units and most gave the answers to an unjustifiable number of decimal places.

Part (iv) was generally competently answered by those who attempted it.

Several answers were fragmentary and some candidates indicated they were short of time.

Question 8

Almost all candidates attempted this question but only a few scored highly. Part (i) was not well done with only a handful of candidates being able to explain the fallacy. Many interpreted the coefficient as a *probability* and said that if it had been 1, then the statement would have been correct! Some of the overseas candidates brought their own preconceptions of age at marriage to the answer rather than basing their responses on the statement in the question.

Many candidates realised that the problem was that the indices were based on 1994 but they did not all use the data to work out the true increase in prices from April 1995 to April 2000 as measured by the index.

Part (iii) was reasonably well answered.

Higher Certificate Paper I – Statistical Theory

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

Several candidates attempted more than the requisite number of questions.

The overall standard was poorer than usual. Results in this paper suggested a continuing decline in candidates' abilities to carry out algebraic manipulations accurately. Analysis involving

(i) the use of disjoint sets,

(ii) moment generating functions,

(iii) conditional probabilities and Bayes' theorem,

(iv) maximum likelihood estimation and Cramer-Rao bound

was particularly weak.

On the positive side, basic manipulations with the Normal distribution are generally satisfactory, as is the interpretation of statistical computer output.

Question 1

With the exception of one perfect answer, all 23 attempts gained fewer than half the marks available.

(i) Most answers were correct.

(ii) Most answers wrongly assumed that $\bar{B}$ and $\bar{C}$ were independent. Very few wrote $\bar{A} \cap \bar{B}$ as the union of the disjoint events $\bar{A} \cap \bar{B} \cap C$ and $\bar{A} \cap \bar{B} \cap \bar{C}$;

(iii) was weak for a similar reason, where $p(B \cap C) = p(A \cap B \cap C) + p(\bar{A} \cap B \cap C)$ is a key step.

(iv),(v) were usually wrongly solved on the basis of independence.

(v) Very few candidates wrote $(A \cap B) \cap (A \cap C) = A \cap B \cap C$.

Question 2

There were few attempts at this question, and most were poor. Candidates may have been put off by the circular table: it may be useful in such a case to place a given man (e.g. the host) in a fixed position and count the arrangements of other people around him. Disappointingly (and surprisingly) there were no good attempts at the Bayesian medical diagnosis problem.

Question 3

This was a popular question, with 51% of answers gaining more than half-marks. Part (a) was generally well done, but (b) was less good. Candidates were seldom able to deal with the mixture distribution (represented by a randomly chosen component) on the basis of first principles. Many assumed that the lifetime was a single Normal variate with a pooled mean (correct) and a 'pooled variance' (which is incorrect). Very few represented $p(X > 2600)$ as

$$p(X > 2600|standard) \times p(standard) \quad + \quad p(X > 2600|high) \times p(high).$$

Attempts to apply the central limit theorem in (b)(iii) were usually on the right lines.

Question 4

This question was also popular, with 57% of passable attempts. However, in many cases the bookwork algebra was 'wishful' or 'fudged': very little use was made of generating functions or of the simple derivation via the sum of independent Bernoulli trial random variables.

(i) Part (a) was well done, but part (b) was poor, the correct answer $1 - [1 - 0.75^{16}]^{12}$ being very rarely obtained.

(ii) Several candidates found the correct $N(-4, 7)$ approximation and then neglected to use a continuity correction.

Part (iii) was a source of great confusion, as many candidates confused numbers of students with numbers of questions and answered in terms of proportions of questions correct, rather than scores in the whole test.

Question 5

This slightly less popular question was poorly done. Graphical work was usually good, and the recurrence relation for $p(x + 1)$ was generally correct. It is clear, however, that most candidates are unable to derive a moment generating function and several were unclear how to use it once found. Very few correct deductions of Poisson$(n\lambda)$ from a correct mgf were seen. In the final part, several candidates confused the question with the mean of 50 Poisson$(1/2)$ random variables, and the correct continuity correction in the Normal approximation so the intended Poisson(25) distribution was rarely seen.

Question 6

40% of the 25 attempts at this question gained more than half marks, weak algebra in constructing the likelihood and differentiating it being a major cause of loss of marks. Very few candidates bothered to check that $\frac{d^2 lnL}{dp^2}$ was negative at $p = \hat{p}$, and there were few sensible attempts at $E(\Sigma X_i)$ (despite $E(X)$being given!). In the numerical part, several candidates used the sample variance of the data rather than substituting their $\hat{p}$ into the given (Cramer-Rao) formula. Some of those who used the correct formula went on to divide its square root by $\sqrt{n}$, so obtaining a confidence interval $1/\sqrt{n}$ times as wide as it should be. No candidates made the comment clearly intended – that sixes were significantly rarer than they should be if the die were fair.

Question 7

There were few attempts at this question, but six of the 11 were good. Again, weak algebra lost several candidates marks, there being several fudges of the given formulae $[F(x)]^n$ and $[1 - F(x)]^n$ and a lack of ability to manipulate as required to obtain the median of X.

Only a few answers obtained the Pareto$(n\alpha)$ distribution for the sample minimum, and several of those obtaining the condition $n > 7.27$ as an equality subsequently rounded down to $n = 7$ rather than up to $n = 8$.

Question 8

Nine of the 16 attempts at this question gained more than half marks. In (i), only a few candidates referred to a 'faster than linear' trend; more saw the endpoints of the graph as 'outliers'. R^2 was well understood, but some candidates used t_{11} instead of t_{10} and introduced a spurious $1/\sqrt{n}$ when calculating the confidence interval for slope. Part (d) revealed a common misinterpretation of a confidence interval as if it were fixed rather than random.

In Part (ii), the graphical comparison was done well but (b) was weak with correct exponentiation being rare $(estimated\ profit = 0.3027 \times 1.0416^{OpCapcty})$. There were few correct comparisons of the confidence intervals in (c).

Finally, in (iii) the log model was usually correctly preferred due to its higher R^2. However, only rare references were made to its better residual plot and its inability to predict a loss.

Higher Certificate Paper II – Statistical Methods

The aim of the Statistical Methods paper is to test candidates' understanding of the fundamental concepts of statistical analysis. The questions require candidates to solve problems involving estimation and hypothesis testing. Particular emphasis is placed upon assessing candidates' ability to summarise and interpret the results of statistical analyses.

In general candidates demonstrate an adequate grasp of the basic techniques required when performing a range of statistical tests and are good at calculating basic descriptive statistics. As in previous years, despite the rubric for the examination stating that 'when a calculator is used the method of calculation should be stated in full', marks continue to be lost by many candidates failing to show the associated working when stating the numerical values of means, standard deviations and so on, obtained from the statistical functions of their calculators.

Matters which caused greatest difficulty included explaining the meaning and uses of statistical tests in general terms, listing or explaining the assumptions required for procedures to be valid, and correctly interpreting the results of statistical analyses. There appeared to be fewer candidates than in recent years choosing to omit these parts of the questions entirely; however, many of the interpretations were incorrect, vague, very muddled or confused.

Candidates should be encouraged to read the questions more carefully. Many continue to lose marks by not actually answering the question asked Whether this is due to carelessness or an inability to understand what is required is not always clear. In addition many candidates waste time by including additional information not asked for in the question.

Other weaknesses are summarised as follows.

1. Graphical presentation of data is untidy and poorly presented. Graph paper is not always used, axes not labelled and titles omitted.

2. Candidates are particularly poor at selecting the appropriate statistical test to perform if this is not stated in the question.
3. Frequently candidates fail to state the null and alternative hypotheses when performing hypothesis tests. Many candidates conclude a question stating that the null hypothesis may be accepted or rejected without having stated what this is.
4. Many candidates are confused between one-sided and two-sided tests. Some candidates state a two-sided alternative hypothesis and then proceed to perform a one-sided test and vice-versa.
5. Many candidates failed to give the values obtained from statistical tables. Many included statements such as 'this test statistic is greater than (or less than) the value in the tables' without stating precisely what the tabulated value was.

Question 1

(i) In the main the statements of the central limit theorem were poor with many candidates having difficulty in clearly explaining the key points. Those candidates who were able to quote a 'textbook' definition often appeared not to understand its meaning.

(ii) Not all the working was given. Many candidates quoted the formula for the 95% confidence interval with the point of the Normal distribution already substituted for a numeric value. This is not sufficient. Some candidates calculated a pooled sample variance and appeared to be confused by differences between the small sample and large sample tests for the difference between two means and the equal and unequal variance assumption.

(iii) Very few candidates gave a correct statement of the formula to construct the 95% confidence interval for the difference in the two proportions. Many could establish that there was a link between the formula and the Normal approximation to the binomial distribution but were unable to follow this through appropriately.

Question 2

(i) Generally candidates were able to make the link between the experimental situation and a binomial experiment but were not always able to explain fully all the necessary assumptions and how they might be assumed to be satisfied here.

(ii) This was generally well answered. Some candidates lost marks by failing to give the null and alternative hypotheses, not explaining that cells with expected counts less than 5 need to be combined when calculating the test statistic, using the incorrect number of degrees of freedom, or misunderstanding when to reject or accept the null hypothesis.

Question 3

The least popular question on the paper; the attempts made were generally poor. Possibly the question was a little long and candidates found it difficult to complete all the parts in the time available.

(i) Many dot-plots were untidy and poorly labelled and many were carelessly drawn with dots not being placed at the appropriate values. Some candidates were unsure what to draw when more than one subject shared the same value.

(ii) This part of the question was answered better than others. Most candidates correctly recognised that they needed to perform a Mann-Whitney U test. Unfortunately some candidates failed to read the question and performed a two-tailed test while others lost marks by not stating the null and alternative hypotheses. Most candidates chose to perform the test using the Normal approximation method. As the tables provided do give values for the Mann-Whitney U test for the subject numbers in the question this was not necessary.

It is arguable how large each sample has to be before the approximation may be used. Textbooks do differ on this. Provided that candidates stated that the samples were large, so that the Normal approximation could be used (e.g. n_1 and $n_2 > 10$), full marks were awarded for correct answers. In general candidates using the standard method rather than the Normal approximation were more successful.

(iii) Answers here were surprisingly poor. Many candidates failed to recognise that they should perform a two-sample t test and many others lost marks by omitting the null and alternative hypotheses, calculating means and standard deviations from statistical functions of the calculator without showing the working involved, using the incorrect number of degrees of freedom or by stating that the test statistic was larger (or smaller) than the tabulated value without giving the value obtained from tables.

(iv) Very few candidates attempted this part of the question, but generally those who had understood the previous parts of the question gave the correct answer.

Question 4

Both parts of this question were generally well answered by those candidates who understood the concepts being examined. However, many candidates failed to include statements of the null and alternative hypotheses in their answers and many had difficulty in explaining the conclusions precisely and stating the assumptions necessary for the analysis to be valid. In addition in (ii) some candidates incorrectly performed a two-sided test with the alternative hypothesis as $\sigma_1 \neq \sigma_2$.

Question 5

(a) This was not well answered, generally because few candidates were able to understand what was required and perform the appropriate sign test. A common mistake was for candidates to perform a Mann-Whitney U test using the judge number to rank the data for those preferring A and B.

(b)(i) The majority of candidates made reasonable attempts at this part of the question, as they were told in the question what test to perform. Again, failing to state null and alternative hypotheses lost marks and frequently when these were given they were imprecise or incorrect.

(b)(ii) Only a few candidates were correctly able to explain that the difference in conclusions was due to the increased power involved in utilising the fact that the data are matched in McNemar's χ^2 test.

Question 6

(i) Histograms were often very untidy and graph paper not always used. As in previous years many candidates did not include titles and although axes were labelled these were often incorrect, especially on the y axis, which many candidates labelled as 'Number of houseflies'. This occurred on scripts from candidates who correctly understood that in a histogram the area of each 'bar' represents the frequency of the group and not the height. Unfortunately many candidates continue to misunderstand this and continue to represent the frequency of each group by the height of the 'bars'.

(ii) Not all the working was given. Many candidates quoted the formula for the 95% confidence interval with the point of the Normal or t distribution already substituted for a numeric value. This is not sufficient.

Question 7

(i) The statement of the model was well done but many candidates failed to state the necessary assumptions for the analysis to be valid – that the terms in the model are additive and that the observations are sampled from Normal distributions with equal variances.

(ii) Candidates can correctly construct a one-way ANOVA table and test for a difference between the groups. Having established a difference it was hoped that candidates would continue to comment on which woodland(s) have soil with superior water holding capacity by examining the mean water holding capacity per woodland. For example pairwise differences between the means could be obtained. Many candidates lost marks unnecessarily by not including the null and alternative hypotheses, the numbers of degrees of freedom or the tabulated value obtained from F tables used to test the null hypothesis.

Question 8

(i) Some candidates failed to read the question properly and produced separate box-plots for the practical and written parts of the examination rather than for the total mark. The box-plots drawn were often untidy, not all included a scale and some were not drawn on graph paper.

Almost all candidates stated that the endpoints of the whiskers are located at the minimum and maximum values in the data set. Whilst this is not incorrect, it represents only one approach. A more meaningful representation of the distribution of the data is achieved by an alternative approach in which the maximum possible length of each whisker is given by 1.5 times the inter-quartile range; any outlier (that is, a value in the dataset outside this range in either direction) is represented by a dot at the appropriate value.

(ii) This was generally well done by those candidates who understood what was required. Marks were lost by candidates failing to show the full working involved in the calculations and by failing to interpret the correlation coefficients obtained correctly.

Higher Certificate Paper III – Statistical Applications and Practice

The syllabus for this paper states that the aim is to develop skills in data analysis using the theoretical concepts developed in the syllabuses for the Ordinary Certificate and Papers I and II of the Higher Certificate to analyse real data sets and communicate the results comprehensively. The objective of the paper is to test these skills.

The candidates' general examination strategy seemed to be to attempt the questions they were most comfortable with first rather than follow the sequence order.

Three candidates attempted more than five questions.

The overall standard was disappointing. There was a general lack of understanding of concepts. The mathematical derivations and calculations were reasonably well done.

Question 1 (22 attempts)
Most candidates appeared to be ignorant of the relationship between the paired t value and the F value.

Question 2 (21 attempts)
There was a lack of understanding as to what interaction is, i.e. an extra effect over that of the sum of the main effects.

Question 3 (25 attempts)
This question was a somewhat changed version of that from a previous paper. It was poorly answered, calling into question the candidates' understanding of the answers to previous years' papers.

Question 4 (13 attempts)
The introduction to this question where an explanation of circumstances was required was poorly answered. In contrast the mathematical derivation was well done.

Question 5 (11 attempts)
Parts (iv) and (v) were poorly answered. Again they tested understanding of concepts. An understanding of overdispersion was generally lacking.

Question 6 (17 attempts)
The best answered question, at least with regard to parts (i) and (ii). It was a question with a strong focus on mathematical derivations and very close in form to a previous year's question.

Question 7 (20 attempts)
Although this was another question very close in form to that of a previous year, it was very poorly answered. The fact that no mathematical derivations and only a small amount of calculation was required could explain this.

Question 8 (13 attempts)
Rather a bimodal distribution of scores. There was a suggestion that some candidates had not been taught this material.

Graduate Diploma Paper: Statistical Theory And Methods I

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.

Overall, the standard of attempts at this year's paper was very good, better than in recent years. Last year's report highlighted three particular areas of concern. At that time, it seemed that candidates were being handicapped by poor mathematical ability; this year, there was no evidence of this problem. Unlike last year, candidates this year seemed much better prepared to tackle standard examples involving standard probability distributions. Finally, while candidates last year seemed unable to exploit links between parts of the same question, this year's candidates were alert to these possibilities.

Question 1
This examined joint, marginal and conditional distributions of two discrete random variables, and was attempted by about two-thirds of candidates. Answers to this question were poorer than those to any

other question on the paper. Candidates had particular problems with part (ii), which required them to find $P(X = x|X + Y = z)$, for z fixed. Many candidates were unable to write down the correct form for this conditional probability in terms of $P(X = x)$, $P(Y = z - x)$ and $P(X + Y = z)$.

Question 2
This tested candidates' knowledge of the Law of Total Probability and Bayes' Theorem (with an example of the Normal approximation to the binomial at the end). About half of all candidates attempted it. The standard of their answers was very variable, with some candidates making excellent attempts while others could hardly get started on the question.

Question 3
Candidates had to derive the mean and variance of the gamma distribution, then work with a bivariate distribution whose marginal distributions were gamma. The general standard was excellent.

Question 4
Candidates were tested on their knowledge of transformations of two random variables. Fewer than half the candidates attempted this question, but the standard of their attempts was good.

Question 5
This tested moment generating functions and the central limit theorem in the context of the binomial distribution. Virtually every candidate attempted this question, and their answers were generally good.

Question 6
This question examined a variety of material: probability generating functions; joint, conditional and marginal distributions; iterated expectation and variance. About half the candidates attempted this question, with variable success. Some candidates went wrong right at the start, when they misspecified the basic random variable as a standard geometric distribution.

Question 7
The question was about simulation, using the inverse c.d.f. method. About two-thirds of candidates attempted this questions, and they all gave very good answers.

Question 8
This tested work on Markov chains, but just one candidate made a serious attempt at it.

Graduate Diploma Paper: Statistical Theory and Methods II

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

Questions 1-7 were the most popular with these being answered by at least half of the candidates. Of these, Questions 2 and 5 were answered well by at least two candidates. Question 8 was not popular.

Question 1
There were only three good attempts at part (i). Although several candidates answered parts (ii) and (iii), none was completely correct. Part (iv) was not answered correctly by anyone.

Question 2
Part (i) was generally well done. In parts (ii) and (iv), few candidates could answer the efficiency and consistency questions. Several candidates considered $Y = max_i(X_i)$ in part (iii).

Question 3
Part (i) was generally well done, but there was only one good solution to part (ii). No candidate correctly derived the critical region in part (iii) and there were only two reasonable attempts at part (iv).

Question 4
Only one candidate gave a correct answer to the bookwork. In part (i), candidates could not complete the square for the posterior distribution. There was only one good solution to part (ii) and one poor attempt at part (iii).

Question 5
Part (i) was generally well done, but few candidates could properly define $\lambda(x)$ in part (ii). No candidate progressed beyond defining the critical region in part (iii).

Question 6
Only one candidate evaluated the stopping boundaries in part (i). Part (ii) was answered well by one candidate and there were only two reasonably good solutions to part (iii).

Question 7
The bookwork was generally well done, but part (a) was not answered correctly by anyone. In part (b)(i), candidates could not properly construct the confidence interval. There were no completely correct solutions to part (b)(ii).

Question 8
This was not a popular question. There was one quite good solution.

Graduate Diploma Paper: Applied Statistics I

Objectives are to test comprehension of theory and methodology applied to a range of problems.

Candidates were generally able to replicate bookwork related to a question but were very weak at demonstrating an understanding of how the theory should be applied. Very often candidates answered a more general question than that asked, failing to engage with the actual data or with the problem being analysed. Further, descriptions tended to be too concise with inadequate attention paid to detail. Some scripts were untidy and difficult to follow.

In preparing for this exam candidates must study analysis not just theory. They must practise describing data and thinking about the types of variables used.

Question 1
Candidates were able to quote the relevant theory but not able to apply it to this problem. This betrays a fundamental lack of conceptual understanding. Candidates need to make sure they understand principles of parameterisation in general linear models.

Question 2
Candidates could describe the method of backward elimination, at least in general terms. Few had the confidence to demonstrate how it operates in practice. Candidates could define influential statistics, but had little idea what to do when influential observations were identified.

Question 3
The point of the question was that, in generalised linear modelling, the deviance alone is not sufficient to define a 'good model'.Few candidates were able correctly to interpret the deviance. Description and interpretation of diagnostic plots was generally poor.

Question 4
Again, few candidates could describe the residual plots or interpret them. This is fundamental to applied statistics.

Question 5
There was a lack of attention to detail on this question.

Question 6
The mathematics in (a) was generally accurate. Answers to (b), which required thought rather than memory, were less convincing. Candidates must learn how to interpret the basic tools of analysis for the range of methods covered in the paper.

Question 7
Few candidates identified the correct factorial model with possible interaction effect. Consequently many analyses were incorrect. Some were difficult to follow. Candidates need to be methodical. Interpretation was generally very poor. It is not sufficient to quote a p-value and to say that an effect is significant. For practical purposes it is important to assess the nature and size of any effect. Here the nature of the interaction should be explored – preferably graphically. There was far too much reliance on p-values.

Question 8
A key feature of this question is that some variables measure time and others distance. Few candidates appeared to have thought about the variables. The nature of the variables influences the way in which one interprets the principal components. Too few candidates knew how to do a scree plot. Clustering on the raw values is dangerous, because of the possible dominance of a few variables; few candidates identified this. Descriptions of the plots were simplistic and lacked an overall grasp of their purpose.

Graduate Diploma Paper: Applied Statistics Paper II

The Applied Statistics Paper II syllabus covers the application of statistical methods to censuses, surveys and designed experiments, and some elementary topics in demography. A total of 23 candidates registered for and sat the paper.

Overall, the performances of candidates on this were poor, and only about a quarter of all candidates were considered worthy of a pass mark. Most candidates had limited knowledge in a few areas of the syllabus prohibiting them from answering all parts of questions chosen. Nine candidates attained fewer than 30 marks, and these candidates really would require considerably better preparation to have a reasonable chance of passing the paper.

On the positive side, six candidates gained marks of 50 or more. Of these, two candidates gained very good marks of 79 and 70. General strengths of candidates in experimental design included: the analysis of data i.e. construction of the analysis of variance for block designs and Latin squares, and other designs such as nested and split plot designs not covered in the syllabus, assumptions underlying

linear models, and how these can be tested. Weaknesses included: significance testing i.e. comparisons of treatment means, partitioning of sums of squares i.e. in the testing of interactions or lack of fit, and interpretation of results. As in previous years' papers, candidates avoided answering questions on response surface design and analysis.

Fewer candidates answered the sample survey questions. Candidates were familiar with analysing sample survey data i.e. estimation of means, variances, and totals based on simple or stratified random sampling. Candidates had limited knowledge of the techniques available for managing practical problems that arise in planning and conducting of sample surveys e.g. non-response, and we would encourage them to read more widely on the topic.

Candidates answered the correct number of questions, and adhered to the rubric.

Question 1 (18 attempts)
Candidates were required to identify the experimental design in section (b). Answers included: nested and split plot designs (neither examined in the current syllabus), and a $4 \times 3 \times 2$ factorial design. Only a few candidates correctly identified the design as a 3×2 factorial design in 4 blocks. Candidates should be familiar with the basic concepts of experimental design i.e. treatment factors, blocking, replication and randomisation. Candidates analysed the data in part (b) using an analysis appropriate to their answer part (i). Often the $temperature \times heat\ time$ interaction was omitted from the analysis of variance, and in some scripts a note added that this could be assessed via plots. Candidates should be familiar with more formal testing procedures for assessing the significance of interactions. Plots should be used to aid the interpretation of results.

Question 2 (13 attempts)
Part (ii) asked candidates to write down a set of meaningful orthogonal contrasts that assessed 5 types of treatment differences among 8 treatments. Most candidates defined *five* contrasts, omitting the quadratic effect of sulphur and the interaction between application timing and the quadratic effect of sulphur. Overall, candidates were less familiar with constructing contrasts that assessed quantitative effects among treatments i.e. linear or quadratic response to increasing sulphur levels.

Question 3 (23 attempts)
Inadequate testing procedures were used to identify pairs of treatments whose effects could be considered significant in parts (iii) and (v). Only a few candidates applied a multiple testing procedure such as the least significant difference. Most candidates were aware that a variance stabilising transformation was required in part (iv) but were less clear why a square root transformation would be appropriate. Why not a log transformation? Candidates should be familiar with the assumptions underlying certain types of data. Only a few candidates commented that the data were counts, and may be Poisson distributed with variance proportional to the mean.

A few candidates misread the question in part (iv), and re-analysed the data using a log transformation whereas the question asked them to transform the data for treatment A using a square root transformation. Thus, time was lost.

Candidates were familiar with the assumptions underlying linear models analysis but less clear how such assumptions, if violated, would impact on the results of the analysis of variance i.e. significance levels and sensitivity of the F and t tests.

Question 4 (9 attempts)
This question required knowledge of 2^k factorial designs and their application as first-order designs in response surface methodology. Marks ranged from 0-8, with one candidate gaining 16 marks.

Candidates could not explain how Normal probability graph paper works with respect to a plot of the effect estimates and interactions from a 2^k factorial design. This is a technique candidates should be familiar with in analysing data from unreplicated factorial designs where there are no degrees of freedom for error.

Three candidates performed the necessary calculations to test for lack of fit. Only one candidate suggested an appropriate design for a 2^k factorial design plus four centre points in 2 blocks of 7 units. Other candidates did not attempt these parts of the question.

To answer part (iii) several candidates re-analysed the data using least squares to fit a first-order model to the data. The regression coefficients could be obtained directly from the effect estimates given in the question without the need for further calculations.

Question 5 (14 attempts)
Examples of sample surveys in which face-to-face interviews, telephone interviews and postal questionnaires would be used were often omitted in section (a). Instead, the disadvantages and advantages of using each method of data collection were listed. Some interesting answers were given in section (b) though refusal of respondents to answer sensitive or personal questions was seldom associated with non-response error. Approximately 50% of candidates were familiar with randomised response methods although some could not explain the principle.

A few candidates gave alternative wording for the questionnaire questions in section (c) without thinking about other information that could be collected to check on the reliability of the respondent's answer.

Question 6 (11 attempts)
This question required a basic knowledge of the calculation of means, variances, totals and proportions for frequency data. One candidate gained 17 marks, but others' marks ranged from 0 to only 4.

Candidates did not know how to calculate basic descriptive statistics for data presented as a frequency table, and could not therefore answer the question.

The data represented a simple random sample of 1 in 20 households. Most candidates were unable to use this information to calculate N, the total number of households in the town. Often $n(=500)$, the number of households in the sample, was used to calculate a point estimate of the total number of cars in the town's households in part (a).

Question 7 (17 attempts)
In parts (iii) and (iv) there was some confusion in the construction of 95% confidence intervals, whether the percentage points of the Normal or t distribution should be used.

Part (iv) was often answered incorrectly since candidates used the estimated population standard deviation, given as 7.75 in the question, as an estimate of the standard deviation of the simple random sample mean.

Candidates were not familiar with the basic considerations in the construction of strata.

Question 8 (5 attempts)
Candidates who attempted this question were familiar with the construction of life tables, and its use in estimating expected age distributions, expected age at death and life expectancy.

Graduate Diploma Option: Statistics for Economics

This section of the options paper is not intended to concentrate on candidates' technical ability and powers of mathematical manipulation. Rather, it seeks to test candidates' understanding of familiar methods and the principles underlying them, and their ability to apply them in economic contexts to tackle real problems with real data that may not have the features assumed by theoreticians.

Candidates should use the mark allocations printed on the papers as an indication of the relative importance of various sections in the paper. For example, it should be clear that answering only the first two parts of question 2 would not lead to a clear pass.

Question A1
The coefficients of c in equations A and C are estimates of elasticities of M with respect to c. Equations B and C do not give elasticities with respect to time, but exponential growth of imports.

Equation D can be usefully re-written $M = -10.2496 + (1.7086 + 0.000217t)c$, showing no exponential growth but slowly rising elasticity.

In part (i) one must use $\mathrm{Var}(X - Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) - 2\mathrm{Cov}(X, Y)$, so the test statistic is $\sqrt{(0.3569)^2 + (0.5851)^2}$. With the original data, an approach using dummy variables would be possible.

Question A2
The worst-answered question. The constant variance hypothesis should be rejected by an F test, but some thought about one-tailed and two-tailed tests was called for.

Inferences about the means of populations from which small samples are taken depend on Normality and homoscedasticity - though the latter assumption can be relaxed by using an approximate test, as exemplified by Minitab.

Candidates appeared to be unaware that a one-way analysis of variance is identical to the usual two-sample t test with pooled variance; note that the output gives `Pooled StDev = 3.91` and `Pooled StDev = 3.906`; also `p = 0.87` and `p = 0.869`.

Question A3
Despite indifferent time charts and poor discussion, this was the best-answered question.

The 'standard errors of the coefficients in the regression' refer to $\hat{\alpha}$ and $\hat{\beta}$ in $I = \hat{\alpha} + \hat{\beta}t$, NOT to the standard deviations of I and t.

Most time charts omitted the source of the data, and many also used unlabelled axes and lacked headings.

The regression line superimposed on the time chart should have provided the basis for an intelligent analysis of UK (under)investment in the context of the Lawson Boom, explained in the first paragraph of the question but ignored by most candidates.

Question A4
The words 'and explain' were ignored by many candidates. The Laspeyres and Paasche price indexes can be usefully explained and understood as the ratios of the costs of baskets of commodities *which must be precisely specified.*

The Laspeyres index is the answer to the third and fourth parts of the question, but justifications were required.

For 1998 the Paasche price index $(1995 = 100)$ was

$$100\frac{\Sigma p_i q_i}{\Sigma p_0 q_i} = 100 \times \frac{224.2}{241.1} = 93.0,$$

The Laspeyres quantity index $(1995 = 100)$ was

$$100\frac{\Sigma p_0 q_i}{\Sigma p_0 q_0} = 100 \times \frac{241.1}{202.4} = 119.1.$$

Graduate Diploma Option: Econometrics

The number of candidates was small, and a meaningful report could not be compiled.

Graduate Diploma Option: Operational Research

The overall standard is fairly poor, although one candidate did reasonably well.

Question C1 (Critical path analysis)
Answers to the parts about network analysis and calculation of mean and SD of activity durations were satisfactory. There was little understanding of PERT assumptions.

Question C2 (LP)
Simplex method appeared to be understood, marred by arithmetic slips. Transportation problem solved very well.

Question C3 (Queuing theory)
Not attempted by any candidate.

Question C4 (EOQ)
Extremely simple question poorly answered on the whole. EOQ formula seemed to be memorised with little understanding.

Graduate Diploma Option: Medical Statistics

The number of candidates was small, and a meaningful report could not be compiled.

Graduate Diploma Option: Biometry

Question E1
This should have been a standard factorial analysis with one quantitative (4 level) and one qualitative (2 level) factor. Only one answer showed any understanding of single-degree-of-freedom orthogonal constants. Diagrams, as always, were poor.

Question E2
A split-plot was recognised by some, but the importance of interaction and testing of means when there is one, are not well understood.

Question E3
One candidate answered part (a), and seemed to have some practical experience.

Surprisingly, (b) was not even chosen as a part-question to make up a few marks, for which purpose it was probably the easiest on the paper!

Question E4
Three candidates answered this. All got something from it – the only question for which this was true!

Explanations of a tolerance *distribution* ignored, for the most part, the idea of a variation in subjects' reaction levels, even if explaining the meaning of tolerance.

Some fair-sized jumps occurred in the theoretical explanations (i) and (ii). Some, but not all, answers contained sketches of probits and logits; but even those which did so were not good on the comparison.

There was one good answer to this question – the only answer substantially above half marks on the whole paper.

Arithmetic in the final part was a bit unreliable.

Graduate Diploma Option: Statistics for Industry and Quality Improvement

The objective of the paper is to allow candidates to show that they have the ability to use statistical methods to improve the performance of industrial processes. They will need a good general knowledge of statistics, together with a clear understanding of specific techniques and how they should be applied in practice. These techniques include statistical quality control charts and related statistics, the design of experiments for improving the product and its manufacture, the use of statistical distributions to model component reliability, and simple probability models for system reliability.

Question F1
Candidates were asked to set up standard Shewhart mean and range charts. The question explicitly referred to the tare weight; that is, the weight of a bottle, which when subtracted from the gross weight gives the (net) weight of the contents. This practical aspect of quality control is not always discussed in text books, but direct measurement of the volume of viscous liquids would be difficult and quite unnecessary. The standard deviation of an individual measurement of volume is increased by less than 4%, from 0.94 to 0.973. The mean falls above the upper action line at sample number 4, and the range is above the upper warning line on the R-chart. In part (iii), the ARL reduces from 500 when the process is on target to 1.26 when the mean is at 111 ml. Since 111 ml is below the lower action line the ARL must be less than 2.

Question F2
This question concerned components of variance at three levels. Sufficient information was given for the components of variance to be calculated without explicitly drawing up an ANOVA table, but it was quite acceptable to do so. The within batch (estimated standard deviation 2.18) and between batch (estimated standard deviation 2.00) variation dominate the variation between deliveries (estimated

standard deviation 0.83), and the standard deviation of a single sample is 3.07. In part (ii) the batch number is a fixed effect and it is paired over the deliveries, which are therefore blocks. The 90% confidence interval for the difference in mean yields excludes 0, so there is some evidence that Reactor B gives the higher yield.

Question F3

This began with an analysis that ignored the possibility of an interaction, and which led to a conclusion that there was no evidence of any effects from varying temperature and humidity. The more appropriate analysis of part (ii) provides some evidence for an interaction, although with only 4 degrees of freedom it does not reach statistical significance, and a statistically significant (at the 10% level) humidity effect. The decision to ignore the length of time for which the process is run is somewhat dubious, because there is some indication of an interaction between time and humidity. It would have been more satisfactory to replicate the experiment and have a reasonable number of error degrees of freedom for investigating interactions. It would also be useful to monitor the variance of the measurements of thickness of the four gloves from each batch.

Question F4

The question was straightforward if candidates had covered this part of the syllabus.