CONTENTS

FOREWORD

This booklet contains reports written by Examiners on the work of candidates in certain papers. **Its contents are primarily for the information of the subject teachers concerned**.

STATISTICS

GCE Ordinary Level

Paper 4040/01

Paper 1

General comments

As in previous years the general level of work on this paper was high. Compared to last year, there was a slightly higher proportion of marks available on this paper for answers requiring comment and explanation rather than graphical work and calculation. In Statistics it is important that candidates should both learn methods and be able to explain why the methods they are using are appropriate.

One very pleasing aspect of the work of almost all candidates was the high standard of graphical work, together with the use of the scales stated in the questions. However, something which both Centres and candidates must address is the need to give final answers to the level of accuracy required by a question. There were numerous examples of a request such as 'calculate to 1 decimal place.....' being repeatedly totally ignored throughout a particular script.

Where a question requires use of graph paper, only the graphical part of the answer should be given on the graph paper. The rest of the solution should be presented in the candidate's script.

Comments on specific questions

Section A

Question 1

Almost all candidates gave correct answers to the first three parts, and a majority to the entire question. A common error in part **(iv)** was a denominator of 32, and in part **(v)** hardly any gave any indication of their working, so that if the result was incorrect, a method mark could not in general be awarded.

Answers: **(i)** 40; **(ii)** 32; **(iii)** 4; **(iv)** $\frac{13}{92}$; **(v)** $\frac{8}{38}$.

Question 2

In part **(i)** far more answers referred to a scenario of 'pieces of paper in a pot' than to, for example, the use of random number tables. The most common error was a description of systematic sampling. For the second mark, the words 'at random' or 'randomly' were required, and these were given by no more than half the candidates. Answers to part **(ii)** were generally very good indeed, but a majority of candidates had difficulty expressing their ideas in part **(iii)**.

Answers: **(ii)** 40, 20, 15.

Question 3

This was generally a very well-answered question. The two most common errors in part **(i)** were failure to round the angles to the nearest degree, and only to consider three sectors. The latter of these also led to a loss of marks in part **(ii)**. The quality of the diagrams presented was excellent. In part **(iii)** only a small minority of candidates failed to relate their calculations to the squares of the radii. Many lost the final mark though through not giving their answer to the requested level of accuracy.

Answers: **(i)** 162, 130, 43, 25; **(iii)** 5.2 cm.

Question 4

Although some tolerance was allowed in readings from the printed graph, there were many answers indicating that sufficient care had not been taken. This was particularly true in part **(iii)**. The answers presented for part **(iv)** suggested that in many cases the words 'at least' had not been interpreted correctly.

Answers: **(i)** 36.5 – 37 years; **(ii)** 13 – 14 years; **(iii)** 16; **(iv)** $\frac{46}{50}$.

Question 5

Most candidates answered part **(i)** correctly, but part **(ii)** was probably the least-well-answered pa… entire paper. In part **(ii)** many just gave one product, not appreciating that both the case when M… arrived late for work, and when he did not, had to be considered. Only a very small number indeed reali… that there was a method of solution which used their answer to part **(i)**.

Answers: **(i)** $\frac{8}{25}$; **(ii)** $\frac{11}{25}$.

Question 6

A majority of candidates scored full marks, the most common errors occurring, as might have been expected, in the calculation of the standard deviation. Very few, having calculated the variance correctly, forgot to take its square root.

Answers: **(i)** 37°; **(ii)** 36.5°; **(iii)** 36.25° or 36.3°, 1.64°.

Section B

Question 7

This was by far the least popular question in ***Section B***, but better candidates who attempted it tended to score very well indeed. Among such candidates, the most common errors were in part **(b)(ii)** where, even if the 'without replacement method' was handled correctly, the denominators were numerically incorrect, and in part **(c)(iii)** where frequently only two products, or three products but containing only two factors each, were given.

Answers: **(a)(i)** $\frac{18}{40}$, **(ii)** $\frac{14}{40}$, **(iii)** $\frac{2}{99}$; **(b)(i)** $\frac{8}{25}$, **(ii)** $\frac{47}{150}$; **(c)(i)** $\frac{7}{10}$, **(ii)** $\frac{3}{40}$, **(iii)** $\frac{53}{800}$.

Question 8

This was an extremely popular question; very few candidates did not attempt it. The quality of the graphs presented was extremely high. There were few examples of candidates not knowing how to calculate the three averages and use them to draw the line of best fit. In part **(v)**, however, there were many examples of a variety of errors, usually related to candidates not obeying the question as regards the instruction 'use your graph' and/or giving the values of a and b to the required level of accuracy. It was not uncommon to see values from the table of data, rather than values read from the graph, being used to estimate the gradient of the line. A surprisingly large number of candidates either drew an incorrect line for part **(vi)**, or totally omitted the final two parts.

Answers: **(iii)** 104 – 105; **(iv)** 81 – 82; **(v)** $y = 1.8x + 32$; **(vii)** 46 – 47.

Question 9

Quite a number of candidates did not read the question sufficiently carefully and in part **(i)** spent time obtaining a single estimate value for the median rather than just stating the median class. A few also made a similar error in part **(ii)** for the mode. Many knew how to approach part **(iii)** but were unable to obtain the correct fraction of the fourth class correctly. Candidates should be aware that when a histogram is being marked, the method mark, on which the other marks depend, is awarded for a clear indication that, in the diagram, area is proportional to frequency. In many of the diagrams presented, this was either partially or totally not the case. It is also necessary for the vertical axis to be appropriately labelled. This can be achieved either by a reference to the scale, or alternatively the words 'frequency density' will suffice. It was pleasing to see the number of candidates who correctly determined the class mid-points for use in parts **(v)** and **(vi)**. Once again in part **(vii)** many candidates lost marks through not obeying the instruction in the question, 'making clear which group of workers you are referring to'. Large numbers of comments could not be awarded any marks because of their ambiguity.

Answers: **(i)** $40 \leqslant t < 55$; **(ii)** $40 \leqslant t < 55$; **(iii)** 52; **(v)** 44.4 minutes; **(vi)** 13.0 minutes.

Question 10

As with most graphical questions, this one was very popular. Two pleasing things in general about cumulative frequency graphs presented were that almost all candidates plotted cumulative frequencies correctly at class upper limits, and almost all also drew the requested polygon rather than a curve. A very common error, however, was to start the polygon from the origin, even though the lower limit of the lowest class was clearly given. In part **(ii)** the vast majority of candidates obtained the number of women correctly, but hardly any were then able to express this as a proportion. The drawing of the graph was usually carried out with considerable accuracy, but the same could not be said for the reading of values from the graph in part **(iii)**. Many marks were needlessly lost in this part. Comments given in part **(v)** were almost universally poor, the concept of range appearing to be a mystery to many. The incorrect comment 'in this company men earn more than women' appeared on the vast majority of scripts. In contrast, a majority of candidates knew how to answer part **(vi)**, and many scored two marks on follow-through from their answers to part **(iii)**.

Answers: **(ii)** $\frac{13}{150}$ or $\frac{14}{150}$; **(iii)(a)** \$14 000, **(b)** \$11 000, **(c)** \$19 000 – \$20 000, **(d)** \$8000, **(e)** \$28 000 – \$28 500; **(iv)** \$13 000; **(vi)(a)** \$14 250, **(b)** \$20 000 (or the comment 'unchanged').

Question 11

Many candidates clearly understood both the procedure for calculating crude and standardised death rates, and their interpretation, and so scored well in part **(a)**. There were, however, a number of small errors which occurred frequently and for which marks were lost needlessly. There were two such examples in part **(a)(i)**. Some candidates, having found correctly the number of deaths in each of the four age groups, then failed to total them in this part. A common arithmetic error among those who converted the standard population figures into decimals was to use 0.8 (rather than the correct 0.08) for 8%. In parts **(a)(ii)** and **(a)(iii)** the death rate units, which could be expressed either in words or symbolically, were omitted by many. However, the overwhelming majority knew that a correct answer to part **(a)(iv)** required a comparison of the standardised death rates. Given the generous tolerance allowed in parts **(b)(i)** and **(b)(ii)** many candidates scored full marks for those two parts, but marks were only rarely scored for the remaining parts of **(b)** usually because, even if a correct decision was given, it was not supported by a correct reason. Among those who did manage to score at all, part **(iii)** was answered far more successfully than parts **(iv)** or **(v)**.

Answers: **(a)(i)** 515, **(ii)** 8.6 per thousand, **(iii)** 6 per thousand; **(b)(i)** 21% – 23%, **(ii)** 55% – 59%.

Paper 4040/02

Paper 2

General comments

Marks obtained covered almost all the available range, and even among those scoring the lower marks, presentation of work was generally satisfactory at worst and excellent at best. Performance on questions requiring numerical work was almost always better than on those requiring decisions and reasons.

In ***Section B***, **Questions 8**, **9** and **10** were attempted by the vast majority of candidates, with **Question 11** being the least popular. It was not uncommon to see this question being attempted as the fourth question in the section, the attempt very soon being abandoned and **Question 7** attempted instead.

Three matters stand out as requiring particular mention.

As the specific comment on **Question 7** details, despite the comment made last year, candidates in many Centres clearly still do not know the correct method for taking a systematic sample.

A general cause of loss of marks among many candidates was failure to observe the required level of accuracy in their numerical answers.

Both **Question 10** and **Question 11** contained an instruction, 'copy and complete the following table'. There was evidence that a few candidates may have entered values in the tables on their question paper and then not copied them into their script.

Comments on specific questions

Section A

Question 1

The most commonly scored marks were 4 and 0, the former being scored by those candidates who both appreciated that this was a question about scaling, and knew how scaling affected a mean and standard deviation. A number of candidates, however, spent many lines of working attempting to calculate numerical results which were, of course, not feasible.

Answers: **(i)** (m – 3), s; **(ii)** 2m, 2s.

Question 2

This was a moderately-well answered question. The answers to parts **(ii)** and **(iii)** were frequently reversed. Some candidates wasted time needlessly copying out the definitions.

Answers: **(i)** B; **(ii)** C; **(iii)** A; **(iv)** D.

Question 3

The quality of graphical work was generally very high, with almost all candidates using the requested scales. However, despite the wording of part **(ii)**, very few candidates then used their graphs to determine that a three-point moving average was required. Many simply seemed to be repeating things they had learned about quarters or seasons of the year, rather than looking for a pattern in what they had drawn. Some clearly thought that the fact the data had been collected over a period of 10 years was the crucial piece of information. Most knew the criterion for determining whether moving averages need to be centred, but the scoring of full marks on this question was extremely rare.

Question 4

Many candidates appeared well-rehearsed in answering this type of question, as similar examples had appeared in previous papers. The most common error was the use of incorrect lower class limits. While it is essential that candidates should show their working in answering numerical questions, it is not necessary, as some did here, to quote lengthy formulae before applying them to the data.

Answers: **(i)** 41.6 mm; **(ii)** 7.5 mm.

Question 5

Although the question stated ‘give a clear reason.....’, many candidates did not do so. Of those giving reasons, and therefore qualifying for associated marks, the vast majority answered part **(i)** correctly, and a pleasing number gave a correct argument for deciding that part **(iii)** was false, but part **(ii)** was answered correctly by only a relative few.

Question 6

Only a very few candidates were able to give a correct reason in part **(a)**, but part **(b)** was generally answered very well indeed, with many gaining full marks. The fact that this question was in a slightly different style to questions on this topic in previous papers did not appear to worry many.

Answers: **(b)(i)** 75.2, **(ii)** 4.

Section B

Question 7

Very few candidates knew how to select a systematic sample correctly, and, as last year, the most common error was, having determined the sampling interval, to apply it to the random number table rather than to the sampling frame. Comments in part **(iv)** tended to relate to methods of sampling in general, rather than to the samples selected in parts **(i)**, **(ii)** and **(iii)** as the question required. There were, however, many candidates who were able to select both the simple random sample and the stratified sample correctly.

Answers: **(i)(a)** 36 70 10 85 46 41 35 59 55; **(ii)(a)** 01 10, **(b)** 04, **(c)** 14 24 34 44 54 64 74 84; **(iii)(a)** 4 large, 3 medium, 2 small, **(b)** 14 36 48 09 63 35 81 60 74.

Question 8

Almost all candidates attempted this question, and many scored very highly on it. Probably the two most common sources of loss of marks were in part **(iv)** where many did not realise that the values of the price relative meant that the price had not changed, and part **(viii)** where many just gave a general comment such as 'the weights have changed' rather than giving a suggestion in context as to *why* the weights had changed.

Answers: **(i)** 140.3; **(iii)** $39 000; **(v)** 104, 107, 100; **(vii)** $67 000.

Question 9

This was another very popular question, it not being uncommon to find candidates scoring full marks. The most common general error was the appearance of the number 140 in the denominator of all probabilities. The vast majority of candidates appeared to have experienced no difficulty in interpreting the context of the question. Even among those unable to answer part **(vi)** totally correctly, most dealt satisfactorily with the 'sampling without replacement' nature of the situation.

Answers: **(i)** 0.464; **(ii)** 0.0667; **(iii)** 0.508; **(iv)** 0.314; **(v)** 0.371; **(vi)** 0.301.

Question 10

Almost all candidates extracted the percentages correctly from the bar charts and presenting them in the required tabular form. Very few candidates, however, then scored more than a further three or four marks, often (as was the case in **Question 5**) through failure to give reasons for their decisions. It is not sufficient to give as a reason for a particular decision, 'as can be seen from the table'. A justification in context is required. For example, in part **(iii)**, to justify why B is false, it is necessary either to give the appropriate calculations, or to state that the same percentage of different totals is being found. Another frequent cause of loss of marks was failure to appreciate that the figures given by the bar charts were percentages.

Question 11

This was the least popular question in ***Section B***, and there were very few completely correct answers presented. In part **(a)(i)** many candidates adopted an incorrect approach of assuming the result which it was required to show. Many correct answers to parts **(a)(ii)**, **(b)(i)** and **(b)(ii)** were seen. However, very few realised that in part **(a)(iii)** it was necessary to obtain the expected winnings, and hardly any, despite the wording of the question, appreciated that parts **(b)(iv)** and **(b)(v)** required finding the two highest expectations, not just the two highest probabilities.

Answers: **(a)(iii)** $1; **(b)(iv)** 8, 10, **(v)** $2.49.