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

Paper 4040/01

Paper 1

General comments

There were many more entries for this subject this year and the overall performance was very good with a large number of candidates obtaining very high marks. However, once again, there were a few Centres whose candidates were unable to make much progress and consequently their marks were poor. There was evidence to suggest that some candidates had been ill-prepared in examination techniques and had used incorrect formulae and incorrect methods.

The responses to the routine questions in ***Section A***, apart from **Question 5** were generally of a very good standard with a number of excellent scripts.

Most of the candidates were able to cope with the range of questions in ***Section B***. Many more candidates attempted the full quota of questions in ***Section B*** with only the poorer candidates managing to finish 3 or fewer questions.

It was pleasing to see, as in previous years, that almost all the candidates used the scales given when answering the graph questions but not all candidates labelled the axes. This year only a few candidates answered the questions relating to the graph on the graph paper instead of on the writing paper which was provided.

Comments on specific questions

Section A

Question 1

Most candidates scored full marks. Most errors occurred in part **(iii)** with a number of candidates giving the amounts for Monday and Thursday separately and not the total amount.

Answers: **(i)** 18; **(ii)** 20; **(iii)** 21; **(iv)** $\frac{2}{5}$.

Question 2

This question was answered well by the majority of candidates. The angles were calculated correctly by most candidates and the pie chart was well done with only a small number of candidates failing to label the sectors or making errors in drawing the required angles.

Part **(iii)** was generally well done but, once again, some candidates did not realise that the radii of the circles are proportional to the square root of the total number of workers.

Answers: **(i)** 102°, 96°, 90°, 72°; **(ii)** pie chart; **(iii)** 7.2 cm.

Question 3

This question was answered well by most candidates but some of the weaker candidates used incorrect formulae for the mean and standard deviation.

Answers: **(i)** 3; **(ii)** 3.1 **(iii)** 2.6.

Question 4

This was a clear and straightforward question with the majority of candidates scoring full marks. In part **(ii)(a)** some candidates thought, incorrectly, that exactly 2 television sets meant 2 television sets and 0 radios.

Answers: **(i)** 61; **(ii)(a)** 98, **(b)** 250, **(c)** 12.

Question 5

Many candidates found this question difficult. Only a few candidates realised that in parts **(ii)**, **(iii)** and **(iv)** they should have used the method of non-replacement. Even fewer candidates realised that in part **(iii)** they could have obtained the answer by using P(at least one white plug) = 1 – P(neither plug is white).

In part **(iv)**, although some candidates used the method of non-replacement they did not multiply by the correct number of arrangements.

Answers: **(i)** $\frac{1}{4}$; **(ii)** $\frac{6}{11}$; **(iii)** $\frac{5}{11}$; **(iv)** $\frac{3}{11}$.

Question 6

There were many fully correct solutions.

In part **(ii)(b)** the weaker candidates often failed to use half of the total number of children from the 16 - 20 class plus the total number of children from the 20 - 24 class.

In part **(iii)** some candidates used an incorrect formula and some candidates used the upper class boundaries for each class instead of the mid-points.

Answers: **(i)** $12 \leqslant t < 16$; **(ii)(a)** 13, **(b)** 26; **(iii)** 14.2 minutes.

Section B

Question 7

This was a popular question with a good number of candidates scoring full marks and many other candidates scoring well.

(a) The majority of candidates correctly calculated the crude death rate of Sandborough and the death rates of the three age groups in Sandborough. Once again some candidates failed to give the respective rate as ..... “per thousand”. When finding the answer for the standardised death rate, some candidates, as in previous years, used the product of the number of deaths and the standard population for each age group instead of the product of the death rate and the standard population for each age group.

Some candidates made hard work of part **(iv)** by first finding the population of Seabourne using the given percentage of population for each age range and then finding the number of deaths for each age range. The straightforward method was to use the product of the death rate and the respective percentage population for Seabourne.

(b) Although there were many correct solutions some candidates failed to draw contiguous rectangles. Other candidates drew the correct heights on their histogram or drew heights in the correct ratio but in some cases their values did not relate to the variable on the vertical axis. Most candidates drew their histogram with equal widths but some candidates incorrectly labelled the end-points of each rectangle.

Answers: **(a)(i)** 16.5 per 1000, **(ii)** 9.5 per 1000, 8 per 1000, 45 per 1000, **(iii)** 21.25 per 1000, **(iv)** 21.68 per 1000, **(v)** 29.4 per 1000, **(vi)** Sandborough, lower standardised death rate; **(b)** histogram.

Question 8

This was a popular question with many candidates scoring high marks.

(i) Most candidates stated the modal class for the length of rods using the correct inequalities but some candidates stated the answer as 80 - 85.

(ii) Some candidates had obviously filled in the table on the Question Paper instead of copying the table on to their script and then completing it.

(iii) A few candidates drew a frequency polygon instead of a cumulative frequency polygon, a few other candidates drew a cumulative frequency curve. A number of candidates plotted the cumulative frequencies against the mid-points of the classes or the lower class boundaries. A large number of candidates failed to label the vertical scale correctly.

(iv) Most candidates used their graph correctly to find the number of rods which measured 72 cm and 88 cm but some did not go on to find the number which measured between 72 cm and 88 cm or even the required probability.

(v) This was quite well done by most candidates but some candidates failed to score marks by not reading the horizontal scale correctly.

(vi) A large number of candidates simply stated that the interquartile range is the difference between the upper quartile and the lower quartile; no reference was made to it measuring the spread of the middle 50% of the distribution.

Answers: **(i)** $80 \leqslant L < 85$;

(ii)

Upper class boundary	65	70	75	80	85	90	95	100
Cumulative frequency	2	8	26	74	138	184	198	200

(iii) Cumulative frequency polygon; **(iv)** $\frac{150 \pm 2}{200}$; **(v)(a)** 82 ± 0.5 cm, **(b)** UQ 86.5 ± 0.5 cm, LQ 77.5 ± 0.5 cm, **(c)** 9 ± 1 cm; **(vi)** A measure of the spread of the middle 50% of the distribution.

Question 9

This was a popular question with many candidates scoring at least 14 marks.

(i) Most candidates correctly plotted the given data using the given scales but, once again, some candidates failed to label the axes.

(ii)(iii) The majority of candidates correctly calculated $(\bar{x}, \bar{y})$ and the semi-averages and most candidates plotted them correctly, so obtaining the correct line of best fit. However some candidates failed to plot the semi-averages and then drew their line of best fit by eye.

(iv) Some candidates gave the wrong students who did not do as well in Paper 2, often stating those candidates who had performed equally as well in both Papers i.e. those on the line of best fit.

(v) A number of candidates gave their answer as a decimal instead of giving it to the nearest whole number.

(vi) This proved to be the most difficult part of the question. Most candidates found the gradient of the line correctly but then assumed that this was the value of a. They also assumed that c was the intercept on the y-axis and tried to find a value for b, often approximating it to 1.

The correct method was to use the calculated points on the line of best fit to find the values of m and c using the equation $y = mx + c$. Then the coefficients of x, y and c should be converted to whole numbers and the equation rearranged in the form $ax + by + c = 0$.

Answers: **(i)** Scatter diagram; **(ii)** mean = (50, 63), lower semi-average = (28, 46), upper semi-average = (72, 80); **(iii)** line of best fit using two of the points from **(ii)**; **(iv)** any two from A, B, D, H or G; **(v)** 61 or 62 marks; **(vi)** $17x - 22y + 536 = 0$.

Question 10

This was not a popular question with very few candidates scoring high marks. This was the fourth ***Section B*** question for many candidates and, in most cases, there was insufficient time left to finish the question.

(a)(i)(ii) This was generally done successfully but some candidates gave their answers as fractions instead of whole numbers.

(b)(i) Most candidates attempting this question obtained the correct answer.

(c)(i) This was done correctly by most candidates but some candidates thought that the order of going into the shops was important.

(ii) Most candidates simply added 0.2 and 0.6. The correct probability is P(she goes into shop *A* and does not go into shop *B*) + P(she goes into shop *B* and does not go into shop *A*).

(iii) The correct probability is P(only one shop and late) + P(both shops and late).

(iv) Expected number of days late = answer to **(iii)** x 200. The answer should be given as a whole number.

Answers: **(a)(i)** 4, **(ii)** 8; **(b)** 11; **(c)(i)** 0.12, **(ii)** 0.56, **(iii)** 0.0516, **(iv)** 10 or 11 days.

Question 11

Most candidates attempted this question and usually managed to score most of the marks available in parts **(ii)**, **(iii)** and **(iv)** but not all candidates were able to draw correctly the histogram in part **(i)**.

(i) This was not answered well. Many candidates drew the correct heights on their histogram or drew heights in the correct ratio but in a lot of cases their values did not relate to the variable on the vertical axis. Too many candidates labelled their vertical axis with frequency or number of days rather than number of days per 10 kilometre or frequency density. Some candidates drew their histogram with equal widths ignoring the groups given in the question.

(ii) Most candidates correctly stated the modal class of the distribution after first drawing the histogram but some candidates obtained their answer, incorrectly, from the table provided.

(iii) Almost all candidates correctly found the mid-points of the five groups of distances travelled.

(iv) Most of the candidates knew how to find the mean and standard deviation and these were often calculated accurately. Some candidates used the end-values for each group rather than the mid-values and others, having found the mean correctly, went on to find the square of each respective deviation from the mean but then failed to multiply by the corresponding frequency.

Answers: **(i)** Histogram with heights in the ratio 6 : 14 : 8 : 2 : 0.8; **(ii)** $30 \leqslant D < 40$; **(iii)** 15, 35, 50, 80, 125; **(iii)(a)** 45 km/day, **(b)** 29.8 km/day.

Paper 4040/02

Paper 2

General comments

In ***Section B*** only a small minority even attempted **Question 11**, and of those who did, many abandoned the attempt very quickly and answered another question instead. It should be noted that where a question or part of a question states that a particular level of accuracy is required for any numerical results, then unless that level is given, the final accuracy mark for that part will be lost.

Comments on specific questions

Section A

Question 1

The vast majority of candidates clearly knew the correct method and were able to apply it. This was one question where the required level of accuracy was almost always given correctly, possibly because the point was specifically mentioned in the report on a similar question in last year's Paper. Only a very small number of candidates attempted a graphical solution, for which a lower number of marks were awarded as the question stated 'calculate'.

Answers: **(i)** 27.2 (litres); **(ii)** 32.8 – 20.5 = 12.3 (litres).

Question 2

Candidates tended either to know precisely what was required to answer this question, or to give a totally irrelevant answer. In part **(i)** correct examples related on the whole to the rolling of dice or tossing of coins, although a variety of others were seen. Rather than giving examples of independent events, some candidates described trials or processes which would produce independent events as outcomes, and such answers were awarded one mark. Apart from examples which were clearly not independent, the most common error was giving probabilities as events. It was permissible, however, for two independent events to be correctly inserted into a probability statement of the form given in the question. The most common error in part **(ii)** was that the events were stated to be mutually exclusive.

Question 3

There were very few correct answers to part **(i)**; the inclusion of the words 'in year D' clearly led a majority of candidates to misinterpret the question. Many more however were able to answer part **(ii)** correctly, and even more to give a comment in answer to part **(iii)** sufficiently detailed to indicate that they had appreciated the point being looked for. A large number of candidates applied the correct method to a total of 25 rather than the correct one of 100 in part **(iv)**. This could not be counted as a misread as, almost always, 25 had been used correctly earlier in the question. The method mark was, however, allowed to score in such circumstances.

Answers: **(i)** $\frac{1}{4}$; **(ii)** $\frac{25}{112}$; **(iv)** 22, 24, 26, 28.

Question 4

In part **(i)** most candidates were able to identify *D* as non-discrete, but hardly any recognised *E* as a constant, even though any comment which implied that this was the case was permitted. When distinguishing between discrete and continuous variables, it is not sufficient just to state that 'discrete variables are counted whereas continuous variables are measured'. It was pleasing to see that many candidates knew that discrete variables are not necessarily restricted to having integer values. Part **(ii)** was generally answered well, follow-through being permitted for *D* if it had not been stated in part **(i)**, but not for *E* as that was not a variable.

Question 5

Many candidates scored well on part **(i)** although answers often were of considerable length. The most concise method for a question such as this is to consider values as being 'so many standard deviations' above or below the mean. Despite the instruction in the question to 'fill in the four blank spaces', (and the given information that each student took only two languages), a number of candidates attempted to give scaled values for the three cells containing a dash instead of, or in addition to, the four blank ones. For part **(ii)** candidates split almost equally into three groups, those who knew exactly what was required, those whose answer clearly showed that they did not, and those who omitted the part totally.

Answers: **(i)** AS 70, BS 76, CF 78, CG 68; **(ii)** Original marks A highest 149, scaled marks C highest 146.

Question 6

A moderately-well attempted question, being another of the type where candidates tended to score either well or poorly. In part **(i)** the reasons given tended either to be all correct or all incorrect.

Section B

Question 7

In part **(i)** most candidates were able to identify 1995 as the 'base year', but not all then referred to either prices or price relatives in their explanation. In part **(ii)** only a minority clearly stated that the prices in 1995 and 2002 were being referred to. Part **(iii)** was the first place in the Paper where even strong candidates often lost marks through not giving their results correct to the required 2 decimal places. Most candidates managed to score at least some marks in part **(iv)**, with many scoring full marks, the method required here clearly being well-known. In contrast, however, completely correct answers to part **(v)** were extremely rare. A few candidates were able to obtain one of the two correct sums of products, but then divided it by 75. A noticeable number gave as their answer to part **(v)** an exact repetition of their answer to part **(iv)**, not appearing to appreciate that it was unlikely that further marks would be being awarded for duplicate work. The mark in part **(vi)** was awarded for any meaningful attempt to compare the final answers to the two previous parts, provided these were not the result of identical working.

Answers: **(iii)** 103.64, 94.34, 108.33, 96.92; **(iv)** 101.3; **(v)** 101.2.

Question 8

Almost all candidates were able to score the mark in part **(i)**. In the rest of the question, even the strongest candidates frequently lost marks through not giving their results to the required level of accuracy. Where no working was shown, a correct result in fractional form was able to score the method mark by implication, but throughout the question, unless a result was given correct to three decimal places, the final accuracy mark for that part could not score. Not surprisingly, parts **(iv)**, **(v)** and **(vi)** were the least well answered. Many candidates, though, were obviously completely at ease with the entire question.

Answers: **(ii)** 0.236; **(iii)** 0.205; **(iv)** 0.137; **(v)** 0.064; **(vi)** 0.135; **(vii)** 0.478.

Question 9

Almost all candidates managed to score the first two marks without any difficulty. Many candidates then managed to score at least some marks in parts **(iii)** and **(iv)**; although many others clearly had no idea how to extract the required figures for individual subjects from the given information about combinations of subjects. Surprisingly, some who worked correctly through part **(iii)**, and therefore of course had already partly solved part **(iv)**, then failed to realise that further application of the same method would complete their answer to part **(iv)**. Most candidates presented bar charts for what they believed were the correct figures for individual subjects, but some, despite the wording of the question, drew them for either combinations of subjects, or percentages. Of these two incorrect formats, the first three marks could be scored by the former, but only the first one by the latter. The standard of drawing was, in general, very good. Where, as in part **(vi)**, a question clearly requests comment in the context of the question, candidates must realise that general comments about the type of diagrams they have drawn will not receive credit. Far too often, the comments made did not relate to previous work in the candidate's answer.

Answers: **(i)** F and G 42, F and S 34, G and S 24; **(ii)** F and G 48, F and S 40, G and S 12; **(iii)** 81%; **(iv)** Boys F 175, S 133, G 152, Girls F 150, S 88, G 102.

Question 10

Most candidates scored the first mark, the most common incorrect answer being "Because it has already been done". The vast majority obtained the value of p correctly in part **(ii)**. In part **(iii)** a number of different sets of seven figures could be used to obtain the value of q. As might be expected, the majority used those for which q was the middle one of the seven, and so obtained a result of exactly 1190. Candidates using these figures were therefore, somewhat fortuitously, not faced with the extra work required of those who used other, perfectly valid, sets of seven figures. The question stated that all the amounts of money were rounded to the nearest \$10, and this was therefore also required of the candidate's value of q, as it was such an amount. Graphs were generally very well drawn, hardly any candidates at all using a scale other than those specified in the question. The most common loss of a mark was due to failure to label the vertical axis. Reading of the required value was usually done correctly in part **(vi)**, but then hardly any were able to use it correctly in the way directed by the question. In part **(vii)** only a minority of candidates made the required comments about the trend and seasonal variation.

Answers: **(ii)** 1090 (or 1090.0); **(iii)** 1190; **(vi)** 1090 (from a reading of 1130).

Question 11

As has already been stated, only a small minority of candidates even attempted this question, and of those who did, the vast majority, having reached part **(iii)**, either answered nothing further in the question, crossed out their answer and attempted another question, or gave answers to the remaining parts which were totally incorrect. Part **(i)** was generally answered correctly, but many candidates lost marks in part **(ii)** through not reading the question carefully. The question asked for the number of ways in which each score could be obtained. What many candidates gave was a description of the methods by which each score could be obtained. At this stage many candidates failed to appreciate the number of ways, and therefore the probability, of obtaining a score of zero, although some, having given the incorrect figure in part **(ii)**, were then able to deduce the correct probability from the requirement that the probabilities should sum to 1. Attempts at parts **(iv)** to **(vii)**, where there were any, then tended to be either completely correct or completely incorrect.

Answers: **(i)** Score

Score	0	1	2	3	4	5	6	8	9	10	12	15	18
Number of ways	6	1	2	2	2	1	3	1	1	1	2	1	1
Probability	$\frac{1}{4}$	$\frac{1}{24}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{8}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{24}$	$\frac{1}{12}$	$\frac{1}{24}$	$\frac{1}{24}$

(iv) \$9.20; **(v)** \$9.20; **(vii)** \$5.