

STATISTICS

GCE Ordinary Level

Paper 4040/01

Paper 1

General comments

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

The responses to the routine questions in **Section A**, apart from **Question 6**, were generally of a satisfactory standard with a number of excellent scripts.

Many of the candidates were able to cope with the range of questions in **Section B**. Although many more candidates attempted the full quota of questions in **Section B**, there were some candidates who only managed to finish 3 or fewer questions. It is important that candidates learn to pace themselves when answering examination papers.

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. As in previous years, a number of 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 assuming that the 12 houses could all be chosen again.

Answers: **(i)** 12 houses; **(ii)** $\frac{4}{15}$; **(iii)** $\frac{1}{15}$; **(iv)** 3 houses.

Question 2

This question was answered badly, with some candidates scoring no marks implying that they had no knowledge of sampling procedures. This is fundamental to a sound understanding of statistical data. Several candidates regurgitated standard rules for the types of sampling, but failed to relate these rules to this particular question. In part **(i)**, 'lottery method' was frequently given, and random sampling was mentioned without any further explanation. Part **(ii)** was answered correctly by the majority of candidates.

Answers: **(i)** Each pupil allotted a numbered card (or names on separate cards) and put in a hat. The cards are shaken and 5 picked out at random; **(ii)** Students arranged in order. First student picked out at random and then every sixth student until complete; **(iii)** 2 boys, 3 girls.

Question 3

This question was answered well by nearly all candidates. The explanations for obtaining the mode, mean and median were usually well done, but some candidates failed to mention that the marks should be ordered before picking out the middle value as the median.

Answers: **(i)** median; **(ii)** the middle mark when arranged in order of size; **(iii)** mean; **(iv)** the sum of the marks divided by the number of students; **(v)** mode; **(vi)** the mark with the highest frequency.

Question 4

This question was answered well by most candidates. In part **(i)**, a few candidates had obviously filled in the tally chart on the question paper instead of copying and completing it. The pie chart was well done with only a small number of candidates failing to label the sectors.

Part **(iv)** was well done by a large number of candidates but, as in previous years, some candidates did not realise that the radii of the circles are proportional to the square root of the total number of vehicles.

Answers: **(i)(ii)**

Type of vehicle	Frequency	Sector angle
Car	14	168°
Bus	3	36°
Van	8	96°
Lorry	5	60°

(iii) Pie chart; **(iv)** 5.4cm.

Question 5

Although this was a clear and straightforward question, a large number of candidates estimated their values from a total of 140 students.

Answers: **(i)** 166cm; **(ii)** 18cm; **(iii)** 68 students.

Question 6

The majority of candidates found this question difficult. Very few seemed to understand the significance of the BLANK squares or the importance of the order of the throws.

- (i)** Few candidates realised that by throwing a 3 or a 5 the player would land on a blank and thus the probability is $\frac{2}{6}$, or $\frac{1}{3}$.
- (ii)** After moving to square 4 on the first throw, a blank square would result if the player threw 1, 3 or 5.
- (iii)** Most candidates ignored this part or failed to record all 7 different scores. Often, correct scores were cancelled out by incorrect scores.
- (iv)** Few candidates correctly obtained 4 ways of moving from START to square 16 after exactly 3 turns.

Answers: **(i)** $\frac{1}{3}$; **(ii)** $\frac{1}{2}$; **(iii)** (2, 6, 5) (4, 4, 5) (4, 6, 3) (6, 6, 1) (6, 5, 2) (6, 4, 3) (6, 2, 5); **(iv)** $\frac{1}{54}$.

Section B

Question 7

This was a popular question, with some candidates scoring full marks and many other candidates scoring well.

- (a)** The majority of candidates correctly calculated the values of P , Q and R . These candidates then went on to calculate the crude death rate and the standardised death rate, with some candidates failing 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.
- (b)(i)** This caused problems for a large number of candidates.
- (ii)** Often, the range was given as 1 to 4, or 4 – 1 and even 20 – 3, or 17.
- (iii)** Some candidates had difficulty giving their answers to 2 significant figures.
- (iv)** Several candidates used increase when referring to the changes from 1958 to 1984. Many candidates simply repeated the information given in the question and did not attempt to express what it meant in terms of the numbers of children per family and the variation in children per family.

Answers: **(a)(i)** $P = 11$, $Q = 160$, $R = 35\,000$, **(ii)** 12 per thousand, **(iii)** 11.55 per thousand; **(b)(i)** not a family or would not be part of a class in a school, **(ii)** range = 3, **(iii)** 2.3 children, **(iv)** families have fewer children; there is a smaller variation in children per family.

Question 8

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

- (i)** This part was not answered well. Many candidates gave a median class as the answer. Other candidates calculated the median, but not correctly.
- (ii)** Most candidates correctly stated the modal class.
- (iii)** Many candidates drew the correct heights on their histogram or drew heights in the correct ratio, but in most 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 bolts rather than the number of bolts per 0.01cm or frequency density. Some candidates drew their histogram with equal widths, ignoring the groups given in the question.
- (iv)** Most candidates correctly found the midpoints of the five groups of diameters.
- (v)(vi)** Many of the candidates knew how to find the mean and standard deviation and these were often calculated accurately but, as in previous years, some candidates did not obtain the correct value for the standard deviation because of premature rounding.

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.

- (vii)** Only the more able candidates successfully completed this part. Many candidates did not know how to obtain, from their histogram, the number of bolts falling outside the given tolerance, and subsequently the percentage of bolts not acceptable to the purchaser.

Answers: **(i)** 2.05mm; **(ii)** $2.04 \leq d < 2.05$; **(iii)** histogram with heights in the ratio 8 : 36 : 26 : 22 : 6; **(iv)** 2.025, 2.045, 2.055, 2.065, 2.08; **(v)** 2.05mm; **(vi)** 0.0404mm; **(vii)** 12%.

Question 9

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

- (i)** A number of candidates drew a frequency polygon instead of a cumulative frequency polygon, and other candidates plotted the cumulative frequencies against the midpoints of the classes or the lower class boundaries.
- (ii)** Most candidates used the correct method to find the median, the quartiles and the percentiles.

- (iii) Most candidates used their graph correctly to find the number of employees who worked 54 hours, but some did not go on to find the number who worked more than 54 hours by subtracting their value from 800.
- (iv) This part was quite well done by most candidates, but some candidates failed to score marks by not marking their graphs clearly, or by not reading the horizontal scale correctly.
- (v) This part was well done by most candidates. Some candidates found 35% of 800, but then did not use this value to read off the value of H from the cumulative frequency polygon.
- (vi) Only a few candidates realised that the histogram would be symmetrical.

Answers: (i) cumulative frequencies 50, 200, 600, 750, 800, cumulative frequency polygon; (ii)(a) 40 hours, (b) 42.5 hours, (c) 45 hours, (d) 32 hours, (e) 53 hours; (iii) 65 employees; (iv) 0.6; (v) 41 hours; (vi) symmetrical.

Question 10

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, as in previous years, some candidates failed to label the axes.
- (ii) The majority of candidates correctly calculated (\bar{x}, \bar{y}) and the semi-averages, and most candidates plotted them correctly. However, some candidates had difficulty plotting the semi-averages and then found that their line of best fit did not pass through all their calculated points.
- (iii) Most candidates were able to use their graph correctly to find the mass of food required.
- (iv) Only the better candidates gave a valid reason for not extending the line of best fit to the y-axis.
- (v) Most candidates were able to draw the line required.
- (vi) The majority of candidates realised that the intersection of the two lines was needed. Most gave the correct answer, but some used the wrong axis for their value.
- (vii) This part was not well done. Many candidates gave the brands of dog food, but without a reason for their choice or a spurious reason.

Answers: (i) Scatter diagram; (ii) mean = (25, 450), lower semi-average = (12.5, 275), upper semi-average = (37.5, 625), line of best fit; (iii) 590g; (iv) not correct; a dog with no body mass does not need 100g of food; (v) graph; (vi) 17kg; (vii)(a) Puppygrow – consumes less food than Doggybix, (b) Doggybix – consumes less food than Puppygrow.

Question 11

This was not a popular question, with only a 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) These were generally done successfully.
 - (iii) This part was well done by most candidates, but some candidates used the method of replacement.
 - (iv) Again, some candidates used the method of replacement.
- (b)(i) Most candidates obtained the correct answer.
 - (ii) This part was answered correctly by most candidates, but some candidates used the method of non replacement.
 - (iii) Most candidates used the correct method, but arithmetic errors were evident in some scripts.

- (c) Many of the candidates attempting this part scored full marks. Some candidates decided to use $\pi = 3.142$, or used the calculator button on their calculator. As in previous questions, some candidates ignored the instruction to round up their answers.

Answers: (a)(i) $\frac{1}{2}$, (ii) $\frac{2}{11}$, (iii) $\frac{2}{11}$, (iv) $\frac{2}{15}$; (b)(i) 0.4, (ii) 0.0064, (iii) 0.1039; (c) 0.69.

Paper 4040/02

Paper 2

General comments

The general standard was similar to that of last year, except that very strong candidates found it easier to score maximum marks on **Section B** questions, rather than losing one or two marks. Most candidates attempted all of the **Section A** questions, and only the very weak candidates failed to attempt four questions in **Section B**.

In **Section B**, **Question 8** was by far the most popular but **Questions 10** and **11** also proved very popular. It was unusual that **Question 7**, involving expectation, proved more popular than **Question 9**.

Comments on specific questions

Section A

Question 1

Many candidates scored maximum marks for this question and it was pleasing to see that most candidates used $\frac{n}{2}$ rather than the incorrect $\frac{(n+1)}{2}$ which has been all too common in previous years. Common errors were incorrect class boundaries, even though these were defined precisely in the question, and incorrect class widths. A large number of candidates failed to give their answers to the specified level of accuracy. It was also disappointing that many candidates failed to state the units, 'hours' when giving their answers, although they were not penalised for this. The small number of candidates who used a graphical method were penalised for not *calculating* their answers.

Answers: (i) 25.6 hours; (ii) $29.4 - 21 = 8.4$ hours.

Question 2

- (i) Most candidates were able to score at least 1 mark in this part.
- (ii) This part was not answered well, and even candidates who may have been able to carry out the operation of centring correctly could not explain clearly what was involved. Many candidates referred, in some way, to finding the mean of two moving averages, but did not mention that they needed to be consecutive moving averages.
- (iii) Candidates who clearly knew what was required in this part often had difficulty in expressing it. All that was expected was 'when the period of the moving average is even'.

Question 3

The majority of candidates failed to realise the significance of 'mutually exclusive' when answering part (a)(i), although many answered part (a)(ii) correctly. A frequent error was to add the probabilities in part (a)(i), obviously confusing the meaning of 'and' and 'or' in this context. Although many candidates obtained the correct answer in part (b), the presentation often left much to be desired, with figures dotted round the page and no indication of what they represented. A number of candidates seemed to think that two separate answers were required, and others did not seem to notice the word 'followed' in the question.

Answers: (a)(i) 0, (ii) 0.75; (b) $\frac{1}{6}$.

Question 4

- (i) A considerable majority of the candidates were able to earn full marks for this part.
- (ii) Most candidates also obtained the correct answer for this part, although the methods used were somewhat longer than expected. It should have been possible to write down the answer immediately, since Amos's mark in Geography equalled the mean for all students.
- (iii) Although there were many successful attempts at this part, a large number of candidates failed to understand the need to determine Barbara's scaled mark ($138 - 90 = 48$) first. This meant that they were unable to make a real attempt at part (iii).

Answers: (i) Amos 78, Barbara 90; (ii) 60; (iii) 32.

Question 5

It was clear that very few of the candidates understood properly the meaning of 'continuous variable' and 'discrete variable'. Even very good candidates stated that discrete implied integer values only. Weaker candidates gave a variety of incorrect answers, many of which suggested that they thought that one or the other of these variables meant a constant. Parts (ii) and (iii) proved a valuable source of marks to almost all candidates, although it was disappointing to see candidates state that B is neither continuous nor discrete.

Question 6

This question was not well answered. Many candidates made statements which, whilst correct, did not answer the question. They must learn to read this kind of question carefully.

Section B

Question 7

Most candidates who attempted this question earned several marks for their attempts at parts (i) to (iv). However, it was quite a common mistake in part (ii) for candidates to give answers showing a '7' or '8' appearing on the first die. A considerable number of candidates did not understand what was being asked for in parts (iii) and (iv). Two problems occurred in the rest of the question. A number of candidates failed to realise that, when calculating expectation, all the sums of money involved need to be in the same units (i.e. £ or p but not a mixture). The second problem was that many candidates were confused over the difference between 'prize' and 'profit'.

Answers: (i) $\frac{1}{8}$; (ii) (4, 6), (4, 6), (6, 4), (6, 4), (3, 7), (2, 8), (iii) 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14;

$$(iv) P(3) = \frac{1}{48}, P(4) = P(13) = P(14) = \frac{2}{48}, P(5) = P(6) = P(12) = \frac{4}{48},$$

$$P(7) = P(8) = P(9) = P(10) = \frac{6}{48}, P(11) = \frac{5}{48}; (v) 20.2p; (vi) £14.90; (vii) 20p.$$

Question 8

Most candidates scored well on this question and the quality of the graphical work was good, but it was disappointing that a number of marks were lost by a failure to label the graphs correctly. There were a number of candidates who did not read the question carefully, and either failed to give answers to the required accuracy, or found the wrong percentages by combining the figures for the two years either in total, or country by country (e.g. for France expressing 240 and 275 as percentages of 515). The latter resulted in meaningless percentage bar charts and scored very few marks. Many candidates failed to relate their comments in part (iv) to the context of the question and lost marks as a result.

Answers: (ii) 1998 France 27.2, Germany 25.6, UK 21.7, Spain 13.6, Rest of Europe 11.9;
1999 France 29.1, Germany 25.4, UK 25.9, Spain 11.1, Rest of Europe 8.5.

Question 9

This was the least popular question in **Section B**, and although there were some excellent solutions, there were also a considerable number of attempts which earned less than half marks. Part **(i)** produced a range of mistakes and even candidates who correctly calculated the annual costs for maintenance and for fuel could not round them correctly to the nearest £50. Often they did not attempt to round their answers at all. This, in turn, made for difficulties in part **(ii)** and a number of candidates made things worse by finding an incorrect price relative for fuel. Many comments in part **(iv)** related to price changes which had already been accounted for in the calculations.

Answers: **(i)** 490, 540 and 932.4, leading to 500, 550 and 950 and hence 10, 11, 19; **(ii)** 112.6; **(iii)** £2180.

Question 10

The basic graphical work in this question was good and the large number of candidates who knew what they were doing scored well on this question. Others clearly did not understand the context and presented a variety of incorrect methods as attempts to calculate moving averages. Good answers to part **(vi)** were not common, although some did mention the need to allow for deviation from the trend line, or to put back the seasonal variation. Very few commented on the non linear nature of the centred moving averages.

Answers: **(ii)** m.a's 573.25, 579, 611.25, 636.75, 609.75, 628.5, 614.5; centred m.a's 576.125, 595.125, 624, 623.25, 619.125, 621.5.

Question 11

Many candidates scored high marks in this question, but even the majority of the strongest were unable to obtain the correct answer to part **(vii)**. It was clear that some candidates had completed the table on the question paper and then failed to copy it on to their answer sheets. Answers were equally acceptable in fraction or decimal form.

Answers: **(i)**

	A	B	C	D	E	Total
Boys	42	16	9	5	–	72
Girls	28	10	12	–	8	58
Total	70	26	21	5	8	130

(ii) $\frac{1}{5}$; **(iii)** $\frac{1}{10}$; **(iv)** $\frac{14}{559}$; **(v)** $\frac{10}{213}$; **(vi)** $\frac{63}{130}$; **(vii)** $\frac{10}{13}$.