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

There was an increase in the entry for this syllabus compared to last year and the overall performance was very good with a large number of candidates obtaining very high marks, showing that they had been well prepared for the examination. This year there were fewer Centres whose candidates were unable to make much progress. However, once again, some candidates used incorrect formulas and incorrect methods.

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

The majority of the candidates were able to cope with the range of questions in ***Section B***. Almost all the candidates attempted the full quota of questions in ***Section B*** with only the poorer candidates managing to finish three or fewer questions.

It was pleasing to see, once again, that almost all the candidates used the scales given when answering the graph questions, but still some candidates answered the questions relating to the graph on the graph paper instead of on the writing paper provided.

Comments on specific questions

Section A

Question 1

Most candidates scored full marks. Any errors seen were usually in part **(iii)** with some candidates giving the amounts for Saturday and Friday separately and not the difference.

Answers: **(i)** 60; **(ii)** Saturday; **(iii)** 120; **(iv)** 950; **(v)** $\frac{1}{19}$.

Question 2

There were many fully correct solutions. Both part **(a)(i)** and part **(b)(i)** were well done by most candidates but part **(a)(ii)** and part **(b)(ii)** caused problems for some candidates. Often the factor 2 was missing and some candidates thought that for part **(a)(ii)** the probability was 1 – their answer to part **(i)**, and incorrect numerators and denominators were used in part **(b)(ii)**.

Answers: **(a)(i)** $\frac{9}{100}$, **(ii)** $\frac{21}{50}$; **(b)(i)** $\frac{1}{7}$, **(ii)** $\frac{18}{35}$.

Question 3

In part **(i)** some candidates thought the method would either be biased because all the students would come from the same class or that the sample was too small. In part **(ii)** some candidates left the four numbers as decimals and a few candidates found the respective proportions of five students instead of twenty students. In part **(iii)** many candidates failed to mention that the first student should be chosen by random selection.

Answers: **(ii)** A5, B5, C6, D4.

Question 4

This question was answered well by most candidates but in part **(ii)** some candidates increased $2… ratio 4 : 3 instead of increasing it in the ratio $4^2 : 3^2$. A few candidates used the angle represent… amount spent on gas instead of subtracting the given angle from 360°.

Answers: **(i)** $132; **(ii)** $512; **(iii)** $218.

Question 5

There were many fully correct solutions with only part **(iii)** producing a number of errors with some candidates quoting 4 or 5 as their answer.

Answers: **(i)** 5.1; **(ii)** 4; **(iii)** 4.5; **(iv)** 7.

Question 6

In part **(i)** few candidates found the proportion of women earning more than $320. Most candidates gave their answer as 4, 3 or even 3.5. Others gave their answer as 56, 57 or 56.5. Parts **(ii)** and **(iii)** were generally answered correctly. In part **(iv)** many candidates referred to men doing more work or harder work or referred to the number of men versus the number of women.

Answers: **(i)** $\frac{1}{15}$; **(ii)** $140 – 145; **(iii)** $170 – 175.

Section B

Question 7

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

In part **(i)** 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. In part **(ii)** a few candidates drew a frequency polygon and some candidates drew a cumulative frequency curve. Only a small number of candidates plotted the cumulative frequencies against the mid-points of the classes. Part **(iii)** was quite well done by most candidates but some candidates failed to score marks by not reading the horizontal scale correctly. In part **(iv)** some candidates correctly read their graph but did not subtract their answer from 800. Parts **(v)** and **(vi)** were well done by most candidates but part **(vi)** was either not attempted by some candidates or the wrong ordinates were used.

Answers: **(i)** 120, 260, 580, 710, 800; **(iii)(a)** 37.5 – 38.5, **(b)** 42 – 42.5, **(c)** 45.5 – 46.5, **(d)** 7 – 9; **(iv)** 60; **(v)** 40.5 – 41.5; **(vi)** 0.56.

Question 8

Many candidates attempted this question with varying degrees of success. In part **(i)** most candidates correctly found the mid-points of the five groups of temperature but, as in previous years, some candidates used the end values for each group. Many candidates knew how to find the mean temperature and standard deviation and these were often calculated correctly. Again, as in previous years, some candidates, having found the mean correctly, went on to find the square of each respective deviation from this mean but then failed to multiply by the corresponding frequency. Part **(ii)** was not answered well. Many candidates drew the correct heights on their histogram or drew heights in the correct ratio but their values did not relate to the variable on the vertical axis. Too many candidates labelled their vertical axis with heights or number of days instead of frequency density or number of days per 2°C. Many others used frequency density but the scale on the vertical axis was half the correct scale. In part **(iii)** most candidates correctly stated the modal class of the distribution. Part **(iv)** was poorly done with only a few candidates correctly using two-thirds of the days in the $14 < T$ 17 group plus the number of days in the $17 < T \leq 21$ group. In part **(v)** few candidates scored both marks. Often candidates repeated the information given in the stem of the question instead giving an interpretation of the information.

Answers: **(i)(a)** 12.7°, **(b)** 3.7°; **(iii)** $12 < T$ 14; **(iv)** 26.7%.

Question 9

This was a popular question with many candidates scoring at least 14 marks. In part **(i)** most ca correctly plotted the given data using the given scales but, as in previous years, some candidates fa label the axes. In part **(ii)** the majority of candidates correctly calculated ($\bar{x}$, $\bar{y}$) and the semi-averages a then correctly plotted them correctly, so obtaining the line of best fit. Some candidates failed to plot the semi-averages and drew their line of best fit by eye. Most candidates correctly answered part **(iii)** and correctly found the equation of the line of best fit in part **(iv)**. Some candidates, when drawing the horizontal scale started at 1 km causing them to give an incorrect reading for the intercept on the vertical axis. In part **(v)** very few candidates gave the correct interpretation, most simply saying it was the intercept on the vertical axis.

Answers: **(iii)** 620 – 630 cents; **(iv)** $y = 62x + 250$.

Question 10

This was a popular question with a good number of candidates scoring full marks.

In part **(a)** most candidates correctly calculated the answers to parts **(i)** - **(v)**. Once again, some candidates failed to give the rates in parts **(i)**, **(iv)** and **(v)** as "per thousand". In part **(v)**, as in previous years, a few candidates 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. Most candidates correctly gave Town *R* as the town offering the better chance of a longer life but not all the answers referred to the fact that Town *R* had the lower standardised death rate.

Part **(b)** was not well done by many candidates. Most candidates correctly found the mean score in part **(i)**. In part **(ii)** it was common to find the scores given as 8 and 8 or 10 and 6, with little reference to the range of the scores. In part **(iii)** many candidates gave three numbers whose sum is 24 but, again, no heed was taken of the range of the scores. A number of candidates tried to solve the question by using algebra with two equations with three unknowns!

Answers: **(a)(i)** 11 per thousand, **(ii)** 160, **(iii)** 35 000, **(iv)** 12 per thousand, **(v)** 10.95 per thousand, **(vi)** Town *R*, lower standardized death rate; **(b)(i)** 7, **(ii)** 11, 5, **(iii)** 5, 9, 10 or 6, 7, 11.

Question 11

This was not a popular question but many of the candidates attempting it scored high marks. The weaker candidates generally performed poorly on this question.

Most candidates gave the correct answers for parts **(i)**, **(ii)** and **(iii)** in part **(a)** and for parts **(i)** and **(ii)** in part **(b)**. Generally, candidates either understood what was required in part **(c)** or made a guess at what was required. There were some very good solutions to this part with very clear explanations.

Answers: **(a)(i)** $\frac{3}{4}$, **(ii)** $\frac{7}{15}$, **(iii)** $\frac{13}{24}$; **(b)(i)** 0.32, **(ii)** 0.56; **(c)** $\frac{115}{208}$.

Paper 4040/02

Paper 2

General comments

Very low marks were comparatively rare this year. There was one ***Section B*** question, **Question 11**, in a form which had not previously appeared in this paper, but many candidates nevertheless made very good attempts at least at some parts of it.

The most noticeable general point this year, though, was that there were a small number of areas in which the overwhelming majority of candidates experienced difficulties. These are detailed in the comments on specific questions, and it is strongly recommended that Centres should ensure that they are noted.

Comments on specific questions

Section A

Question 1

Most candidates displayed a very good understanding of the concepts being tested in this question, with the vast majority scoring highly. A few appeared to misunderstand the question, not realising that two answers were required for each part.

Question 2

This question was generally very well answered with almost all candidates obtaining the means correctly although, as might have been expected, the standard deviations caused somewhat more problems. One point which it could be useful for Centres to explain to candidates is that the number of marks available for a part of a question is a good guide as to how much work is required to obtain the correct answer. A few candidates wrote a page or more of calculations in attempting to answer this question, not something which it might be expected would be required when only four marks are available.

Answers: **(i)** 70, 20; **(ii)** 66, 22.

Question 3

Almost all candidates obtained the correct results in parts **(i)** and **(iii)**. Fewer candidates than might have been expected obtained the correct result for the mean, many failing to do so because, although the table stated very clearly what each figure represented, candidates used what were stated to be sums as means. Others divided by something other than the total frequency. There were very few correct answers to part **(iv)**, and this is one of the areas of concern mentioned in the general comments. The errors were partly due to the same reason as the errors in calculating the mean, but more worryingly many candidates did not appear to be aware of the form of the formula for the variance which it was necessary to use to answer this part.

Answers: **(i)** 390; **(ii)** 15.6; **(iii)** 7487; **(iv)** 7.49.

Question 4

The mid-points were found correctly more frequently than were class intervals, with some candidates clearly unaware of what was meant by a class interval. Correct answers to this question are found by using the given information to convert the stated class limits into the true class limits, and it was frequently obvious that this had not been done. Also, although one of the variables was discrete and the other continuous, many candidates treated them in an identical manner, which resulted in very few scoring full marks.

Answers: **(a)(i)** 1749.5, **(ii)** 499; **(b)(i)** 164.75, **(ii)** 10.

Question 5

In order to score marks in both parts of **(i)**, candidates had to give a correct reason before stating their conclusion. Part **(a)** was usually answered better than part **(b)**. Any comment relating to the existence of an intersection, such as that its probability was not 0, was considered sufficient reason, but just a statement that the events could occur together was not, as it is not justifying why they could. Many clearly knew the independence property but lost both marks because they stated that 0.12 (the product of the two separate probabilities) was equal to 0.1 (the probability of the intersection). Two events are only independent if those two values are exactly equal. In part **(b)** Venn diagrams were very rarely seen, yet it was candidates who used that approach who more frequently reached the correct result. Errors in alternative methods were both common and very wide in variety.

Answer: **(ii)** 0.3.

Question 6

This question produced another issue of general concern. Very few candidates indeed realised th not necessary to calculate any standard deviations. Consequently on most scripts there was evidence more calculations having been carried out than need have been the case. Nevertheless, even among th who did provide large quantities of calculation, there was frequently evidence of an understanding of wha was required. The most common error in parts **(ii)** and **(iii)** was for the four individual marks to be summed, rather than a comparison of individual marks being made to achieve the correct answer to part **(iv)**.

Answers: **(i)** 60, 36, 58, 46; **(ii)** 75, 76, 85, 84; **(iii)** 100, 89, 109, 102.

Section B

Question 7

For candidates with a sound knowledge and understanding of probability this question was a plentiful source of marks. Almost all candidates attempted this question, and almost all of those managed to score the first six marks without undue difficulty. Some of the problems in the later parts stemmed in many cases from a misinterpretation of the question, the expressions 'at most' and 'at least' not being understood. Also, although parts **(iv)** and **(v)** related to people, it was quite common to find denominators relating to houses being persevered with.

Answers: **(i)** 25; **(ii)** 65; **(iii)(a)** $\frac{2}{25}$, **(b)** $\frac{7}{25}$; **(iv)(a)** $\frac{18}{65}$, **(b)** $\frac{33}{65}$; **(v)** $\frac{1}{208}$.

Question 8

High marks were not achieved frequently on this question for a variety of reasons. It was also not as popular as might have been expected. Almost all who attempted it were able to score the first two marks however. Part **(ii)** was also usually answered correctly, although often after far more work than was necessary. Many then lost marks in the later parts as a result of not reading the question sufficiently carefully. The correct ordered pairs were often stated in part **(iii)** but not related to the appropriate values of Y, resulting in all the marks for this part being lost. Despite the question stating that the die was biased, and the given probabilities clearly indicating this, many candidates then produced a solution to part **(iv)** based on an assumption that the die was unbiased. Candidates making this error were able to score two marks. One was granted as a 'special case' if the unbiased distribution was completely correct. The other was the final mark, given for presenting results in the form of a table as the question requested. Some candidates realised their 'unbiased error', and corrected it, on obtaining a result other than that given in part **(v)**. Others abandoned their attempt at this question totally at that point. Very few candidates appreciated the 'quick way' of obtaining the answer to part **(v)** using that to part **(i)**.

Answers: **(i)** 2; **(ii)** 2, 3, 4, 5, 6, 7, 8; **(iv)** 2 0.16, 3 0.24, 4 0.25, 5 0.2, 6 0.1, 7 0.04, 8 0.01.

Question 9

The first four marks were scored by almost all candidates. High marks were also common for the graphical work, although it was rare to see an appropriately titled graph. Only a small minority of candidates knew the correct reason in part **(iv)**. Part **(v)** was another of the areas of general concern. Hardly any candidates at all appeared to know that the seasonal components must sum to zero, as the moving average technique is simply smoothing out the seasonal fluctuations, and not increasing or decreasing the overall total of the data. The value of q is therefore obtained by setting the sum of the given components to zero and solving the resulting equation. Most appreciated that in part **(vi)** it was necessary to take a reading from the trend line and then add the appropriate seasonal component to it, although this was not always done correctly.

Answers: **(i)** 284, 279, 67.5, 67.25; **(v)** –6.4.

Question 10

Many candidates scored highly on this very popular question, although full marks were rare, usually as a result of a failure to answer the final part in sufficient detail. Almost all scored full marks in part **(i)** although some did not immediately appreciate the arithmetic required in relation to the balls. Correct price relatives were usually obtained in part **(ii)** although frequently after far more work than was necessary. A price increase of 3%, for example, immediately means a price relative of 103. A few failed to note the word 'decreased', even though it appears in bold print. Parts **(iii)** and **(iv)** were generally well done, with calculation errors being rare. Comments made in answer to questions such as that posed in part **(v)** must be meaningful, in sufficient detail, and in context.

Answers: **(ii)** 103, 90, 102, 105; **(iii)** 101.8; **(iv)** 7125.

Question 11

The use of a random number table in this way to examine knowledge of various methods of sampling had not been examined previously in this amount of detail, but many candidates showed they had a moderately good knowledge of the methods, with one notable exception. Many gave correct reasons in the first two parts of part **(i)** and then selected a correct random sample. However, of the areas mentioned in the general comments which appeared to have been learned correctly only by very few candidates, that of by far the greatest concern is the procedure for selecting a systematic sample. Most candidates knew that it related items at regular intervals, but very few knew how to select the first item correctly. Most seriously of all, however, was the fact that a very small proportion of candidates appeared to be able to apply the sampling interval correctly. Having determined that every tenth item was required, almost all then incorrectly selected every tenth number from the random number table, rather than correctly selecting every tenth item from the population listing. The first item selected was 07, and so the remaining four were then 17, 27, 37 and 47. In contrast, a majority were able to determine correctly the gender ratio for the stratified sample, and to select such a sample totally or almost totally correctly. Many lost marks in part **(iv)** as a result of not reading the question sufficiently carefully. It specifically asked about the three samples selected, yet many candidates simply gave descriptions of the general sampling methods. Knowledge of quota sampling, as displayed in answers to part **(v)**, was very sketchy. It was extremely rare to see mentioned one of the most important properties of the procedure – that a population listing is not required.

Answers: **(i)(c)** 01, 11, 02, 15, 26; **(ii)(a)** 01, **(b)** 10, **(c)** 07, **(d)** 17, 27, 37, 47; **(iii)(a)** 3 men, 2 women, **(b)** 08, 12, 44, 38, 28.