STATISTICS

Paper 4040/01

Paper 1

General comments

The overall performance of candidates continued the trend set last year; there were fewer very high marks than had been the case in previous years, but very few extremely low marks.

It is regrettable to have to report that all the points mentioned under the heading 'General comments' in the report on last year's paper have to be repeated.

Once again numerous marks were lost by candidates who did not give their results to the required level of accuracy stated in a question. There was also a needless loss of marks, albeit less frequent than in the past, by candidates failing to obey correctly the instruction 'copy and complete' in relation to a blank table given in a question. A few candidates are clearly still completing the table on their question paper rather than giving a completed table on their answer paper.

Many candidates are still obtaining relatively fewer marks for the 'comment' parts of answers than for numerical parts purely because they are not reading a question sufficiently carefully, and therefore not giving their comments in adequate detail. This point will be expanded upon in the comments on individual questions, but it continues to be the case that many candidates produce general comments that have obviously been learned by rote from a textbook or similar source, when what is clearly asked for is a specific comment in the context of the question.

The error mentioned in detail in the report on last year's paper relating to the calculation of semi-averages again occurred very frequently. This is something which some Centres particularly need to rectify.

One extra point which needs to be mentioned this year concerns the plotting of points on graphs. Many candidates plotted data by very faint pencil 'dots', some of them scarcely visible. Credit cannot be given to a graph if the points on it cannot be seen. Candidates are recommended to plot data points with small, but clear, crosses, and summary points calculated from the data, such as semi-averages, by dots within a small circle.

Despite the various errors mentioned in this report, however, the overall performance of the majority of candidates remains good, and is to be commended.

Comments on specific questions

Section A

Question 1

Many candidates obtained full marks, the most common error being to give the highest value of the set of data as the mode.

Answer: **(i)** 102; **(ii)** 102; **(iii)** 109.

Question 2

A minority of candidates were unable to use the information given in the question to obtain the value of x, and were therefore unable to complete the rest of the question. The vast majority of those who were able to obtain this value then completed the question completely correctly. Among these candidates, the most common error occurred in part **(iv)** where it was forgotten to include in the total those houses without any residents.

Answer: **(i)** 68; **(ii)** 17; **(iii)** 59; **(iv)** 200; **(v)** 487.

Question 3

This was easily the worst-answered question on the paper. Most obvious from the answers presented was the fact that very few candidates had any knowledge of the process of quota sampling. Knowledge of the use of postal questionnaires and the method of stratified sampling was clearly much better, but in most cases answers tended to be in the form of general definitions, rather than being specific comments relating to the contexts mentioned in the question. To obtain both the available marks in part **(ii)**, for example, a candidate needed to give a general comment about the greater accuracy of stratified sampling, and then give a specific example of that accuracy in the given context, e.g. by using stratified sampling, the ratio of males to females in a sample could be guaranteed to match exactly that in the population.

Question 4

Most candidates scored very well on this question. The most common error was the omission, in the interpretation of the value of x in part **(ii)**, of the fact that the people had not visited France.

Answer: **(i)** 3; **(iii)** 60; **(iv)** 53.

Question 5

As mentioned in the 'General comments', some candidates did not copy the blank table given in **(i)** into their answer book before completing it, and hence lost marks needlessly. The probabilities in parts **(ii)** and **(iii)** caused no problems for most, but only a small minority realised that the given game score in part **(iv)** meant that only two of the possible 16 outcomes needed to be considered.

Answer: **(ii)** $\frac{3}{16}$; **(iii)** $\frac{1}{4}$; **(iv)** $\frac{1}{2}$.

Question 6

The cumulative frequencies were usually calculated correctly and plotted correctly against the vertical axis. However, marks were lost otherwise in relation to the drawing of the graph for a variety of reasons: the requested scales were not used, axes were not annotated (particularly the vertical axis with 'cumulative frequency'), the graph was not started from a point on the horizontal axis, and a curve was drawn rather than the requested polygon. By far the most frequent error, though, was the plotting of incorrect values against the horizontal axis; the correct class limits being 699.5, 709.5 etc. Almost every one of these errors meant that an incorrect value was read from the graph in part **(ii)**.

Answer: **(ii)** 712 hours.

Section B

Question 7

As has been the case in the past, those candidates who had acquired an understanding of this topic were able to score very well on this type of question. Once again, though, even those candidates whose working was perfectly correct often lost marks through not giving the death rates to the requested level of accuracy. In part **(v)**, despite the question referring to 'the same standard population', some candidates used the percentage population for the second town in their calculations. Very few indeed knew that the situation described in part **(vi)** could arise from the two towns having vastly different age-structures in their populations, i.e. far more old people than young in one town and vice versa in the other. A considerable majority clearly knew the criterion which determined the correct answer to part **(vii)**.

Answer: **(ii)** 7.12; **(iii)** 4.57, 7.60, 12.83; **(iv)** 6.36; **(v)** 7.75; **(vii)** *A*, because of its lower standardised death rate.

Question 8

The point made in the 'General comments' regarding the plotting of points on graphs is particularly relevant here.

As always, this was a very popular question, but still one on which many candidates lost marks because of incorrect calculation of the semi-averages. The points used to calculate the lower semi-average, for example, are the five with the lowest x-values, not the first five given in the table. Almost all candidates scored no marks at all for part **(iv)** because they had not read the question sufficiently carefully. The question firstly asked for an interpretation of the point at which the line of best fit met the vertical axis. This is that it is the point which gives the estimated number of sales of the book if no money has been spent on advertising it. Most candidates gave that as their reason why it would be unwise to use the point for estimating sales, the general belief clearly being that if nothing was spent on advertising there would be no sales. The correct reason was that as the x-value of the point, 0, was well below the lowest x-value in the table, estimation would involve extrapolation which is always a dangerous practice. In part **(v)** about half of all candidates drew the line $y = x$ correctly, but very few were able to give the requested comparison using their two lines correctly. Many candidates believed the cost of the book was in some way involved, rather than the correct 'because the first line is above the second for all values of x, advertising was more effective in selling the first book'.

Answer: **(ii)** (340,454), (229,327), (451,581); **(iii)** 520.

Question 9

This was the least popular question in ***Section B***, but was usually answered very well by those who attempted it. The most common error, even among those candidates whose solutions were otherwise based completely correctly on the principle of area being proportional to frequency, occurred in part **(iv)**, where many gave the modal class as that with the tallest rectangle rather than the one with the rectangle of greatest area.

Answer: **(ii)** 7, 13, 20, 5; **(iii)** 2.4; **(iv)** 40 – under 50; **(v)** 1.33.

Question 10

This was a very popular question, in which the first eight marks were scored by a majority of candidates. However, various errors then led to a loss of marks later in the question. In part **(iii)** many candidates did not round their answer to an integer. In part **(v)** some candidates got their answers 'the wrong way round' or did not give the amount by which the median increased. Only a minority of candidates were able to deduce how to use the graph to answer part **(iv)**, and then most of those answered part **(vi)** incorrectly by reading the graph at a value of 66 minutes. Those people who took 56 minutes for their journey on the second day would have taken 46 on the first day, and so that is the value at which the graph should have been read in part **(vi)**.

Answer: **(i)(a)** 44, **(b)** 37, **(c)** 50, **(d)** 28, **(e)** 60; **(ii)** 64–66; **(iii)** 78th percentile;
(iv) $\frac{102}{120} = 0.85$; **(v)(a)** median increases by 10 minutes, **(b)** IQR unaltered;
(vi) $\frac{72}{120} = 0.6$.

Question 11

A large number of candidates only represented the 2004 figure in part **(i)(a)** by two and a half squares rather than the correct three and a half. Correct answers to part **(i)(b)** were very rare indeed. The incorrect answer seen to part **(ii)(a)** in many cases, 'the mean is affected by extreme values', is as good an example as there could be of 'rote-learned' answers being produced rather than correct answers being obtained from the context of a question. In this case the arithmetic mean was totally inappropriate because the data was qualitative, not quantitative. Pie-charts were generally well-drawn, and a pleasing majority of candidates knew that they had to work with the squares of the radii in **(c)** even though some were then unable to do it correctly. Only a small minority of candidates were then able to work through the final two parts correctly.

Answer: **(i)(b)** 1.87 cm; **(ii)(c)** 72, **(d)** 22, **(e)** 10, 20, 10, 10.

STATISTICS

Paper 4040/02

Paper 2

General comments

Unlike last year, good marks were obtained on all ***Section B*** questions by some candidates.

There was, however, a return to the situation of a few years ago when a number of marks were lost by candidates who did not give their results to the required level of accuracy stated in a question. There was also a needless loss of marks, albeit less frequent than in the past, by candidates failing to obey correctly the instruction 'copy and complete' in relation to a blank table given in a question. A few candidates are clearly still completing the table on their question paper rather than giving a completed table in their answer booklet.

Despite the error being referred to quite frequently in previous reports, many candidates are still obtaining relatively fewer marks for the 'comment' parts of answers than for numerical parts purely because they are not reading a question sufficiently carefully, and therefore not giving their comments in adequate detail. This point will be expanded upon in the comments on individual questions, but it continues to be the case that many candidates produce general comments that have obviously been learned by rote from a textbook or similar source, when what is clearly asked for is a specific comment in the context of the question.

Comments on specific questions

Section A

Question 1

As has often been the case in the past, there was still considerable confusion in the minds of many candidates about the concepts of independence and mutual exclusivity. In contrast, in part **(b)**, where just one of the two was being considered, many candidates scored well.

Question 2

Candidates almost always knew exactly what to do and scored highly, or displayed a total lack of appreciation of which formulae to apply. On occasion, though, a candidate whose approach to part **(i)** showed a complete lack of understanding of how to calculate the weighted mean, then used a correct standard deviation formula in part **(ii)**.

Answer: **(i)** 55.3; **(ii)** 16.2.

Question 3

The number of correct answers to parts **(i)** and **(ii)** was pleasing, but correct answers to part **(iii)**, requiring evaluation of the correct proportion of 70, were very rare.

Answer: **(i)** 50.5, 55.5; **(ii)** 5; **(iii)** 36%.

Question 4

This was a question of a type very common on this paper in the past, and almost all candidates knew how to approach it. Probably the most common cause of numerical errors in the final result was premature approximation. Many candidates lost the final mark through not giving their result to the nearest minute as requested by the question.

Answer: 133 minutes.

Question 5

A majority of candidates were able to obtain the first four marks by one of the accepted methods or another. There were very few completely correct answers to the final part, though, because many candidates who did manage to obtain a correct algebraic expression did not express their answer in the required form, i.e. *s*, (or a term in *s*), on its own on one side of an equation.

Answer: **(i)** 130; **(ii)** 38; **(iii)** $s = 32.5 + 1.25r$.

Question 6

This was probably the most successfully answered question on the paper. Only a small number of candidates spent time needlessly evaluating frequencies for all grades.

Section B

Question 7

In general, answers to this question followed the trend of the last few years as regards ***Section B*** probability questions. But this year, although many candidates scored highly, very few obtained full marks, unlike the situation some years ago. This was because many good candidates, who scored the other 15 marks, lost a mark in part **(iv)** through failure to give their answer correctly to 3 significant figures. One particular trend in common with recent years was that quite a number of candidates who experienced difficulties with parts **(iii)** and **(iv)** had no problems with part **(v)**.

Answer: **(ii)(a)** 0.740, **(b)** 0.187, **(c)** 0.840; **(iii)** 0.154; **(iv)** 0.0706; **(v)** 0.376.

Question 8

Many candidates scored all, or almost all, of the eleven marks available in parts **(ii)**, **(iii)** and **(iv)** by giving completely correct solutions to the numerical parts of the question, but scored hardly any marks at all in parts **(i)** and **(v)** where comments were required. Time and again answers given to part **(v)** showed that answers learned 'by rote' were being quoted without thought, as they bore no resemblance or relevance to the context of the question. It was, for example, quite common to see a reference to 'prices', when prices were in no way referred to in the question, and played no part in the context. Another comment seen frequently was ‘the wage rates may have altered’ which was meaningless when what the question asked was why an estimate based on the changed wage rates might be inaccurate. In contrast, comments relating specifically to the context of the question, such as that this supermarket was just one of a number owned by the company, were very rarely seen. It cannot be stressed too strongly that general comments not referring to the context of a question will not score any marks.

Answer: **(ii)** 125.0, 118.2, 118.5, 127.3; **(iii)** 123.5; **(iv)** $3.1 million.

Question 9

This question was answered much more successfully than were ones on the topic only a few years ago, but there are still parts for which answers are frequently incorrect, and this is often Centre-related. Candidates were generally successful in obtaining the required simple random and stratified samples, but many have clearly still not learned the correct procedure for obtaining a systematic sample. Then in part **(iv)** a considerable number totally ignored what the question asked, and gave general descriptions of the three methods of sampling, obtaining no marks for that part as a result. Candidates must answer what a question asks, in this case for a comparison of the accuracy of the three samples actually selected. Many candidates managed to score one mark in part **(v)(b)** through giving an identical criticism for the two different situations, and any valid point could only score for one of them. Answers to part **(v)(a)** very rarely scored any marks, almost always through being too general and irrelevant. There is no point in just stating that a proposed method of sampling 'will be biased'. To score a mark it is necessary to clarify why and in what way it will be biased.

Answer: **(i)(a)** 66, 01, 30, 06, 45, 43, 11, 29, 10, **(b)** 6, 3, 0; **(ii)(a)** 00, 09, **(b)** 01, **(c)** 11, 21, 31, 41, 51, 61, 71, 81; **(iii)(a)** 4, 3, 2, **(b)** Albion 07, 11, 26, 05, Briony 63, 50, 69, Cherry 76, 84.

Question 10

Quite a number of candidates lost marks needlessly in part **(i)(b),** either because they did not give their results to 1 decimal place, or because their approximation to that level of accuracy was incorrect. Where, as was the case here, the calculations involved in an answer are very simple and straightforward, marks will be lost unless they are exactly correct. Despite comments in the report in previous years, some candidates tried to apply some method of their own in part **(iv)** rather than using the required method that seasonal components should sum to zero. In part **(v)** many candidates did not apply the relevant seasonal component to their reading from their graph.

Answer: **(i)(b)** 8.9, 9.4, 9.6, 10.6, 10.7, 10.9, 10.7, 11.2, 11.5, 12.0; **(iv)** –5.7; **(v)** 11.55 thousand (or 11550).

Question 11

This question was both more popular, and answered more successfully, than questions on expectation and games have been in the past. Very few candidates scored full marks, however, because even most of the more able believed that they had answered part **(vi)** fully on obtaining the result \$1.94, failing to realise that it was also necessary to take into account the probability that the other person won the game.

Answer: **(ii)** 0, 1, 2, 3, 4, 5; **(iii)** Probabilities $\frac{6}{36}, \frac{10}{36}, \frac{8}{36}, \frac{6}{36}, \frac{4}{36}, \frac{2}{36}$; **(vi)** \$11.67.