Version : 1.0 0608

ASSESSMENT and QUALIFICATIONS ALLIANCE

General Certificate of Education

Mathematics 6360
Statistics 6380

MS/SS1B Statistics 1B

Report on the Examination

2008 examination - June series

Further copies of this Report are available to download from the AQA Website: www.aqa.org.uk

Set and published by the Assessment and Qualifications Alliance.

The Assessment and Qualifications Alliance (AQA) is a company limited by guarantee registered in England and Wales (company number 3644723) and a registered charity (registered charity number 1073334).
Registered address: AQA, Devas Street, Manchester M15 6EX *Dr Michael Cresswell Director General.*

General

Overall, this paper proved more accessible than recent papers. Consequently, there was both an increase in the proportion of scripts achieving high marks and a reduction in the proportion failing to achieve at least 30 (raw) marks.

The majority of candidates scored well in Questions 1, 2 and 3 (except for part 3(d)) and in parts of Questions 5 and 6. Whilst most candidates were able to score at least minimal marks in Questions 4 and 7, certain parts, particularly in Question 7, proved accessible only to the higher achievers.

Most candidates, as expected, used their calculators' statistical functions to maximum effect in parts (a) of Questions 1 and 3, but often much less so in part (a)(i) of Question 7 due, in the main, to failing to identify correct mid-points or, more seriously, through ignoring the frequencies. The minority of candidates who used their calculators' more advanced statistical functions to simply state answers in Questions 5 and 6 suffered severely when their answers were incorrect. Centres are reminded that examiners are not expected to try to deduce the reasoning behind a stated incorrect answer and also that multiple undeleted answers/solutions generally lose marks since the mark gained is the average for those answers/solutions offered. Many candidates completed the scatter diagram in ink rather than in pencil. Those who then tried to correct mistakes often lost marks for some points being unclear. Finally, there appeared to be an increase in the submission of anonymous inserts that were also not attached to answer books; both these practices should be actively discouraged.

Question 1

For the great majority of candidates, this question proved a positive start and many scored full marks. Use of regression functions on calculators was by far the most common approach in part (a) with the result that most obtained correct values for a and b. A minority of candidates confused a and b, so losing at least 5 of the 6 marks available for the question. There remained a significant number of candidates who spent valuable time using formulae to calculate values for a and b. Whilst most such candidates were successful, a small number calculated b^{-1} and, of course, all used up valuable time. In answering part (b), the majority of candidates substituted $x = 21$ correctly into their equations, but a significant proportion obtained 31.85 from $\left(\frac{33.0+30.7}{2}\right)$ to score 1 mark.

Question 2

This was a very accessible question for candidates of all abilities. Many candidates scored full marks and the awarding of fewer than 5 marks was rare indeed. Candidates who extracted the necessary information from the table generally had more success than those attempting the use of probability formulae. Few candidates failed to score the 3 marks available for parts (a) and (b). In part (c), common errors were to quote answers of $0.4+0.7=1.1$, $0.4\times0.7=0.28$ or $0.4+0.7-0.28=0.82$, which appeared to indicate, as on previous papers, a lack of understanding of the addition law for two (non-mutually exclusive and dependent) events; something that needs attention in future. Many candidates coped well with the conditional probabilities in parts (d) and (e). When marks were lost, it was usually for quoting $\frac{42}{400}$ or $\frac{42}{70}$ in part (d) and/or for quoting $\frac{21}{400}$ or $\frac{21}{120}$ in part (e).

Question 3

Many candidates scored the first 8 marks in this question. Those candidates who used the correlation function on their calculators almost invariably scored full marks in part (a) though a small minority lost 1 or 2 marks through quoting the answer to less than 3 significant figures, such as 0.81 or 0.8. Candidates who calculated r using a formula often did so with good understanding but sometimes less than accurately. In part (b), most candidates recognised that there was 'positive correlation' but some did not attach an adjective to indicate the strength. Almost all candidates included in their statements a reference to the context usually by mentioning 'length' and 'width'. Whilst most scatter diagrams scored full marks, it was not unusual to see the points not labelled or a point, particularly $\mathrm{I}(244, 128)$, plotted incorrectly. Answers to part (d) were generally poor, often scoring no marks. Common incorrect answers were r, $\frac{r}{2}$, a value or values above 0.5, a range of values such as $0 \text{ to } 0.4$, or phrases such as 'no correlation' or 'strong correlation'. Marks were only gained for an appropriate value linked to a source.

Question 4

Despite this type of question appearing on recent papers, the ability of far too many candidates to correctly identify the median and quartiles from a list of values remains disappointing. In part (a), some candidates failed to even order the values, whilst other quoted the positions in the list as their values for the median and quartiles or used the runs scored as frequencies. Nevertheless, many candidates obtained 40 for the median and 50 for the interquartile range or at least 13 and 63 for the quartiles. In part (b)(i), most candidates stated incorrectly that no value occurred more than once whilst others, having stated that the mode was 0, merely repeated the words of the question. Some candidates argued incorrectly that, since a was unknown, the mode could not be found. Answers to part (b)(ii) were better, with most candidates stating correctly that, since a or the largest value was unknown, the range could not be calculated.

Question 5

Many candidates scored between 8 and 12 marks on this question with only the best scoring the full 15 marks. In part (a)(i), almost all candidates realised the need for standardisation and the great majority of them completed it correctly for 3 marks. Again, in part (a)(ii), many candidates obtained the correct final answer though sometimes by unusual routes. The majority of others either introduced their answer to part (a)(i) [perhaps an idea gained from similar questions on previous papers?] or failed to perform the necessary area change and subtraction. A minority suggested that $0.8-(1-0.8)$ indicated a probability of 0.6 or 0.72575. In part (b), many candidates scored 3 marks by showing a correct method but using $z=+1.0364$ instead of $z=-1.0364$. Perhaps a simple sketch would have shown that the value must be less than 140. Fewer candidates than in similar questions on previous papers used $z=0.85$, but there was an increase in the number who simply stated 142.6, presumably from their calculators' inverse normal functions, and so scored 0 marks. Answers to part (c) showed a welcome marked improvement with many candidates finding correctly the variance or standard error of $\bar{X}$ and then standardising correctly. Sadly, many then failed to find the correct area and so lost 2 marks. Again, a simple sketch would have shown that the answer was greater than 0.5.

Question 6

Parts of this question proved a good source of marks for many candidates with the more able scoring all 15 marks. Answers to part (a)(i) were usually correct and found, as was intended, from tables. A minority of candidates calculated $\mathrm{P}(M=15)$ using the formula. As always, part

(a)(ii) caused difficulties to many candidates. Whilst almost all of them attempted a subtraction using values from tables, many considered at least one incorrect value of M. The requirement of 'more than 10 but fewer than 20' translated to $\mathrm{P}(10 < M < 20) = \mathrm{P}(M \leq 19) - \mathrm{P}(M \leq 10)$. Most candidates used the correct formula in part (b), but a significant minority approximated 0.29 by 0.30 so as to use tables, resulting in a loss of all of the 3 marks available. Answers to part (c)(i) showed an improvement in the knowledge of the relevant formulae. A small proportion of candidates apparently chose to ignore the emboldened word '**do**' and used $p = 0.29$ rather than $p = 0.71$, resulting in a loss of at least 4 marks since the comment marks in part (c)(ii) were dependent upon correct answers in part (c)(i). Nevertheless, many candidates were able to score at least 2 of the final 3 marks. When a mark was lost, it was usually for stating that "since the samples were not random, the claim was not justified". A number of candidates made no references to their answers in part (c)(i), but simply stated that "the samples could not be random as they only included women".

Question 7

This question proved to be the most difficult on the paper. It was very disappointing to see so many candidates, some scoring high marks on other questions, failing, either by using their calculators' statistical functions or from first principles using formulae, to find correct values in part (a)(i). All too often, incorrect mid-points were identified or, more seriously, frequencies were ignored. Attempts at part (a)(ii) were often equally disappointing. Whilst some candidates ignored the word 'Hence' and so repeated part (a)(i) with revised mid-points, many of those who tried to deduce answers made a complete hash of the process by apparently not realising that 1 hour was equivalent to 60 minutes. Thus it was all too common to see candidates adding 1, or even 100, to both mean and standard deviation values or multiplying one or both values by 60 or 100. Future candidates clearly need to be much better prepared in these areas of the specification. In answering part (b)(i), most candidates were aware of the relevant formula for a confidence interval but lost 1 accuracy mark due to the aforementioned errors in part (a). Candidates who used an incorrect z-value or, more seriously, used $n = 8$, lost more marks here. In part (b)(ii), the most common acceptable reasons were "actual times unknown" or "mid-points used" whereas the common response of "mean and standard deviation are estimates" did not gain the mark available. The scoring of marks in part (c) was rare indeed. Most candidates failed to identify the difference between Vernon's two claims and so referred to their sample mean values for both. His first claim required reference to 74% (or 26%) from the original data, whereas his second claim required a comparison of 1 hour or 60 minutes with a correspondingly **correct** confidence interval in part (b)(i). A small minority of candidates considered that the claims were best answered by commenting on Vernon's plumbing skills and work ethic.

Mark Ranges and Award of Grades

Grade boundaries and cumulative percentage grades are available on the **Results statistics** page of the AQA Website.