Version 1.0

General Certificate of Education (A-level)
January 2011

Mathematics MS2B

(Specification 6360)

Statistics 2B

Report on the Examination

Further copies of this Report on the Examination are available from: 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.

General

AQA regrets that an error was made in the setting of this paper, which meant that part (a)(ii) of question 4 was invalid: the expected method required a knowledge of linear combinations of random variables that is outside the specification. Consequently, all candidates were awarded the 2 marks available for this part of the question.

This problem notwithstanding, candidates appeared to be very well-prepared for this examination, and it was very pleasing to see many fully correct solutions to each of the questions. However it is still the case that some candidates do not state their hypotheses in the correct way and neither do they form correct conclusions in context; these are often far too positive in nature. The calculation of s when summarised data is given still needs to be addressed, as do the conditions under which z-tests and t-tests are conducted. The wrong test is used far too often, resulting in a loss of marks.

Question 1

In part (a), it was expected that candidates would understand that for a sample of any size from a normal population with known variance, the reference distribution is $\mathrm{N}(0, 1)$. Far too many candidates based their choice of distribution on the small value of n, and consequently incorrectly used t_{n-1} as their reference distribution. This resulted in a loss of all marks for this part of the question.

In part (b)(i), most candidates correctly used a t-distribution. Some candidates still used the z-distribution, with $z_{\text{crit}} = 1.6449$ often seen. Although most candidates correctly found $\overline{x} = 550$, there were far too many who thought that $s = \sqrt{\frac{334}{9}}$ or, even worse, $s = \sqrt{334}$. In part (b)(ii), far too many candidates failed to state the level of significance, even though this was specifically asked for in the question. Others incorrectly thought that the level of significance was 90%. Many candidates did not justify their comments by correctly stating that 545 was not in the confidence interval; incorrect statements, such as '**it** is not in the CI', gained no credit.

Question 2

As in previous series, this proved to be a good source of marks for many candidates. In part (a), it was disappointing to see the many candidates who did not seem to understand the term 'contingency table'. Consequently, the 2 marks available were awarded to those candidates who filled in the table with the expected frequencies or retrospectively for those who calculated the expected frequencies correctly in part (b). Others, who simply copied the values from the table given in the question or who produced values which did not tally with the given totals, gained no credit.

In part (b), many excellent solutions were seen. However, in a few cases, the hypotheses were not stated and z or t values were stated instead of the required χ^2 value of 11.345. Also, rounding errors arising from the calculation of the expected frequencies often resulted in the calculated value of X^2 falling outside the acceptable range.

Question 3

Part (a)(i) was answered very well. However, in part (a)(ii), some candidates attempted to use $\mathrm{Po}(1.5)$ or, because $\mathrm{Po}(0.15)$ was not in the booklet provided, attempted to average the values found under $\mathrm{Po}(0.10)$ and $\mathrm{Po}(0.20)$. Neither of these approaches gained any credit.

In part (b), it was expected that candidates would state either that 'X and Y are independent' or that 'the number of catches and run-outs are independent'. Although these were often stated, there were candidates who thought that stating 'they are independent' or that 'cricket matches are independent' would suffice. These statements were not sufficient to gain any credit.

Parts (c)(i) and (c)(ii) were usually very well answered. However, some candidates seemed to have some difficulty interpreting statements such as 'at least 15'.

Question 4

In part (a)(i), candidates usually managed to find the correct values for $\mathrm{E}(X)$ and $\mathrm{Var}(X)$.

Part (a)(ii) was considered to be invalid, as described at the beginning of this report.

In part (b), there were many excellent solutions seen. Unfortunately, as is usual when the probability distribution is defined using formulae, several candidates thought the distribution to be continuous and consequently used integration methods in their attempts at $\mathrm{E}(T)$ and $\mathrm{Var}(T)$ for no reward.

The majority of candidates wrote down the correct value of 0.8 as their answer to part (c)(i). Part (c)(ii) was only answered well by the most able of candidates, and so the correct answer of 0.09 was very rarely seen. Many candidates attempted to find the correct probabilities with some success. However, they then multiplied their answer by 0.8 (or the value they had obtained in part (c)(i)), which rather negated the good work that they had done up to that point. Probably the most efficient ways of doing this part of the question were either to evaluate $\mathrm{P}(X = 2, 3 \text{ or } 4) \times \mathrm{P}(T = 3 \text{ or } 6) - \mathrm{P}(X = 4 \text{ and } T = 6)$ as $0.8 \times 0.15 - 0.01 \times 0.3 = 0.09$ or to complete a table of the required probabilities as shown below.

	$X = 2$	**$X = 3$**	**$X = 4$**
$T = 3$ (0.05)	$X + T = \mathbf{5}$ $0.05 \times 0.1 = 0.005$	$X + T = \mathbf{6}$ $0.05 \times 0.4 = 0.02$	$X + T = \mathbf{7}$ $0.05 \times 0.3 = 0.015$
$T = 6$ (0.1)	$X + T = \mathbf{8}$ $0.1 \times 0.1 = 0.01$	$X + T = \mathbf{9}$ $0.1 \times 0.4 = 0.04$	$X + T = \mathbf{10}$ $0.1 \times 0.3 = 0.03$

Some candidates included $X = 1$ and consequently gained an answer of 0.12, which some then again multiplied by 0.8 to obtain the very popular incorrect answer of 0.096. Only the most able candidates were able to answer part (c)(iii) correctly.

Question 5

In part (a)(i), most candidates provided the required response. In part (a)(ii), as neither the distribution of the population nor the variance were specified, z-values, 1.6449 (David) and 2.3263 (James), or t-values, 1.660 (David) and 2.364 (James), were permitted.
Although some conclusions were far too positive, most candidates scored very well on this part of the question. Part (a)(iii) was not at all well answered. David and James made use of the Central Limit Theorem because the population was not known **or** not stated as being normal. This was not often seen, even by some very good candidates.

In part (b), most candidates managed to state the correct reasons for the required responses: 'Type I error' (David) and 'no error' (James). Those candidates who chose not to give a reason for their choice lost both marks.

Question 6

The graph in part (a) was usually adequately sketched. There was, however, a minority of candidates who interpreted the request for a 'sketch' as an excuse to offer up a very poor quality freehand drawing. This should not have been the case and any such attempts were likely to be penalised.

The majority of candidates used simple areas of rectangles to obtain their very efficient solutions to parts (b)(i) and (b)(ii). Most candidates realised that, when the answers are given in the question, sufficient working must be shown in order to obtain full credit. Some candidates, when they are confronted by a continuous distribution, seem to think that integration methods always have to be used. This is not the case when straight lines form part of the graph of the probability density function. Those candidates who used valid integration methods usually gained the correct given values and thus full credit.

Candidates were expected to use the information given in part (b) to write down the answers to parts (c)(i) and (c)(ii). Although some candidates recognised that the answers to part (b) implied that the upper quartile of X was $8\frac{1}{3}$ and that the lower quartile of X was 3, many attempted to recalculate these values, usually incorrectly, and mostly by the use of calculus methods.

Disappointingly, in part (d), only the most able candidates recognised that, whatever the values obtained for the median and lower quartile, $\mathrm{P}\left(X < \text{median} \mid X \geq \text{lower quartile}\right) = \frac{1}{3}$ and $\mathrm{P}\left(X < \text{median } \mathbf{and} \ X \geq \text{lower quartile}\right) = \frac{1}{4}$ are always true.

Mark Ranges and Award of Grades

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