Version : 1.0 0609

General Certificate of Education

Mathematics 6360
Statistics 6380

MS/SS1B Statistics 1B

Report on the Examination

2009 examination - June series

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

The level of achievement on this paper, both in terms of the average (raw) mark and the spread of (raw) marks, was in line with that expected and of a similar level to that achieved on recent papers in summer series. Judged by the notation used by candidates and by the presentation of their solutions, most appeared to be prepared to at least an adequate standard. In fact, the proportion of scripts achieving high marks far exceeded the proportion gaining minimal marks. The majority of candidates were able to score well on each of the first four questions. Most were then able to make worthwhile attempts at some, often all, of the final three questions but often with an increasing loss of marks.

In particular, all but the strongest candidates lost most, often all, of the 7 marks available for questions 5(b)(iii), 6(c) and 7(b)(iv). Candidates continued to make good use of their calculators' statistical functions in questions 2, 4 and 5, and appropriate use of tables in questions 3, 6 and 7, rather than giving unsupported answers direct from their calculators' advanced statistical functions. On a less positive note, attention needed to have been drawn, in many cases, to instructions 1 and 6 on the front page of the question paper since it was not unusual to see:

- answers in blue ink or even in pencil;
- graphical work completed, often untidily, in ink rather than in pencil;
- anonymous unattached inserts.

The following general errors often resulted in a loss of marks:

- premature approximation — working to three, or even less, significant figures;
- multiple undeleted answers — for which an average mark, rounded down, was awarded.

Question 1

This relatively standard probability question enabled most candidates to score at least 7, sometimes 9 but rarely 11, marks. Full marks were usually scored in parts (a)(i) and (ii). Many candidates also scored full marks in the remainder of part (a) but it was not unusual to see an answer based upon $\frac{60+32}{160}$ for part (a)(ii) and/or answers based upon $\frac{18}{60}$, $\frac{48}{160}$ or even $\frac{30}{48}$ for part (a)(iv).

Fully correct answers to part (b) were rare. Although many candidates were able to construct the expression $\frac{24\times56\times32}{160^3}$ for 1 mark or the expression $\frac{24\times56\times32}{160\times159\times158}$ for 2 marks, few realised that permutations had to be considered. Of those who did, about half used 3 rather than $3! = 6$. The very small number of candidates who based their correct answer on

$$\frac{\binom{24}{1}\binom{56}{1}\binom{32}{1}}{\binom{160}{3}}$$

are worthy of a special mention!

Question 2

Whilst there remain candidates who calculate the value of r using a formula, the vast majority use their calculators' inbuilt correlation function. Using either method, the correct value of r was usually obtained in part (a), though a few weak candidates stated, here and in question 4, values for a, b and r. Most interpretations in part (b) were in context and correct but a minority of candidates omitted the word ‘positive’. In part (c), almost all candidates plotted the points correctly, but significantly more did not label their plotted points.

Most candidates identified the two most likely female snakes in part (d)(i). A minority of candidates clearly re-calculated the value of r $(= 0.488)$ in part (d)(ii). Those candidates who estimated the value gave quite varying answers: some less than 0.25, others greater than 0.9. Accompanying interpretations were often comparative, using words such as ‘weaker’ or ‘stronger’, whilst others omitted the word ‘positive’.

Question 3

Almost all candidates knew how to standardise, and there were thankfully very few instances of attempted ‘continuity corrections’, something that is not required in this unit. The majority completed part (a)(i) correctly, but too many did not perform the necessary area change. Whilst almost all candidates realised that a difference of areas was required in part (a)(ii), there were varying degrees of success, again often due to not making the necessary area changes — perhaps a prime example of the benefit of sketches.

In answering part (a)(iii), many candidates performed a variety of standardisations and subtractions of areas for 1 mark. Given that the request for $\mathrm{P}(X = x \mid X \sim \text{normal})$ has appeared on previous papers, the answer should have been known. Attempts at part (b) were generally quite good with even weak candidates scoring 2 or 3 marks for equating their standardised expression to a recognisable z-value. The usual error, even by better candidates, was to use $+2.0537$ rather than -2.0537 and then to disregard or even hide the fact that the resulting standard deviation was negative.

Question 4

Many candidates scored full marks on this regression question. Almost all candidates found accurate values for a and b using their calculators’ inbuilt regression functions, although a few quoted b to only two significant figures. Thankfully, very few candidates using calculator functions interchanged the values of a and b. Some very able candidates continued to use a formulae approach, and in almost all cases were successful, but appeared to penalise themselves as regards time available for answering other questions.

Most candidates correctly quoted their value for a or 50 as the answer to part (b), but a small minority substituted $x = 1$ into their equations. Whilst almost all candidates changed 13 weeks to 91 days, somewhat fewer then substituted $x = 91$ into their equations. Those who did were often able to interpret the resulting very small negative value of y as a justification for the claim. Some candidates made alternative or additional points about the justification requiring extrapolation.

Question 5

Candidates appeared better prepared for part (a) than in the past but it remains a concern that far too many candidates cannot sensibly tackle this type of question. Whilst most candidates were able to correctly identify the median, the IQR appeared to be an unknown term or was found using $75 - 25 = 50 \;\Rightarrow$ median $= 3$. A small minority of candidates even based their answers on the cumulative values of $\sum fx$. In part (a)(ii), answers, usually directly from calculators, were in the main correct, although incorrect answers due to working only with x-values or only with f-values were seen.

It was pleasing to note the large proportion of correct answers to parts (b)(i) and (ii). In the former, some candidates mysteriously obtained the correct answer despite totally incorrect work in part (a) and, in the latter, some answers were backed-up by a correct calculation.

In part (b)(iii), only the most able candidates were able to make valid statements as to why a normal distribution was not an appropriate model. Many invalid attempts made reference to the sample size or the Central Limit Theorem, neither of which had any relevance here. Candidates should be aware that, if X is normally distributed, then, for any sample size, the sample mean, $\bar{X}$, will be normally distributed regardless of the Central Limit Theorem; for other cases, the Central Limit Theorem states that, for sufficiently large sample size, the sample mean will be approximately normally distributed, but candidates should not infer from this that the population values, or actual sample values, will themselves be approximately normally distributed.

Question 6

Most candidates knew how to find a confidence interval for μ and, despite somewhat unusual notation at times, many were able to score full marks. Where marks were lost, it was usually for using an incorrect z-value. Answers to part (b) revealed an improvement in candidates' knowledge of the distribution of the sample mean, in particular that $\mathrm{Var}\left(\bar{X}\right)=\dfrac{\sigma^2}{n}$ rather than σ^2 . Such candidates usually scored full marks, but a minority lost a mark for rounding the z-value to 2.7 rather than 2.71 . As already mentioned in question 5, knowledge of when and why the Central Limit Theorem should be used was weak and confused, with many candidates only considering the sample size rather than the fact that the distribution of Y was unknown.

Question 7

Some weak candidates attempted to answer part (a) using a normal approximation, something that is not part of this unit's specification. Virtually all other candidates used the appropriate binomial tables but frequently made one or more errors by finding $\mathrm{P}\left(R\le 10\right)$ in part (a)(i) and/or using $\mathrm{P}\left(R\le 9\right)$ or $\mathrm{P}\left(R\le 5\right)$ in part (a)(ii).

Candidates appeared well prepared in the use of the binomial formula, with almost all scoring full marks in part (b)(i). However, answers to part (b)(ii) were very disappointing. All too often, candidates calculated $\mathrm{P}\left(S=1\right)$ or $1-\mathrm{P}\left(S=1\right)$ rather than $\mathrm{P}\left(S\ge 1\right)=1-\mathrm{P}\left(S=0\right)$, with some even using $n = 22$ rather than 35.

In answering part (b)(iii), most candidates took note of the word 'correctly' and the number '120' in the question to then score both marks. However, a minority lost marks for stating answers of 7.2 or $2.60\left(\sqrt{6.768}\right)$. Part (b)(iv) proved beyond all but the high achievers. Whilst some candidates based their comments on the perceived practical difficulties in sorting letters, others decided that there was only one assumption and so made their comments unclear. Those candidates who did treat the statements as two separate assumptions, still often demonstrated little understanding of the implications of 'equal means' but 'different variances'. Many such candidates used the means to comment on independence and the variances to comment on the probability. As a result, the awarding of full marks for this final part was very rare.

Mark Ranges and Award of Grades

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