Version : 1.0 0608

ASSESSMENT and QUALIFICATIONS ALLIANCE

General Certificate of Education

Mathematics 6360

MS2B Statistics 2B

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

There were many excellent solutions seen to each of the eight questions on this paper. It was pleasing to see that, in general, the majority of candidates had been well-prepared for the examination. However, some candidates are still failing to state hypotheses in the correct way or even not at all which often nullifies any computations carried out and conclusions drawn. In questions where the answer is given, candidates must be careful in showing sufficient working to gain full credit.

Question 1

Although this first question provided most candidates with a good start to the paper, there were some who, in part (a), either failed to state any hypotheses or who thought that the null hypothesis should indicate 'an association', rather than 'no association', between the incidence of asthma and the volume of traffic. Also simply stating "H_0: not associated" and "H_1: associated" was not really sufficient when a statement in the context of the question was required. Although the great majority of candidates realised that Yates' correction had to be used, there were far too many candidates who could not apply it correctly. A conclusion in context was required: simply stating "Reject H_0" was not sufficient to gain full marks. Although most candidates formulated a correct conclusion in context, some either used statements which were too positive in nature or made no statement at all. In part (b), the question asked for a comment about the 52 children who had asthma, given that they lived in an area where the volume of traffic was heavy. It did **not** ask, as many candidates thought, for a comparison with the children who lived in an area where the volume of traffic was light. Consequently, the required response "More than expected had asthma" was not always seen. Some candidates, having drawn the wrong conclusion in part (a) that there was 'no association' between the incidence of asthma and the volume of traffic, still felt justified in making this statement.

Question 2

Part (a) was answered well by most candidates. The use of tables to evaluate $\mathrm{P}(X=8)=\mathrm{P}(X\le 8)-\mathrm{P}(X\le 7)$ or the use of the formula $\mathrm{P}(X=8)=\dfrac{\mathrm{e}^{-6}\times 6^8}{8!}$ were equally successful methods used here. In part (b)(i), although the correct answer of $\lambda=9$ was often seen, it was common to see the wrong answer of $\lambda=3$. In part (b)(ii), the most prevalent error was made by candidates who wrongly thought that $\mathrm{P}(Y>9)=1-\mathrm{P}(Y\le 8)$. In part (c)(i), where candidates were asked for the distribution, it was not sufficient to simply quote the value of λ without also indicating that the distribution was Poisson. In part (c)(ii), the phrase 'at most 20 telephone calls' was sometimes misinterpreted, with candidates wrongly attempting to evaluate $\mathrm{P}(T>20)$ instead of $\mathrm{P}(T\le 20)=0.917$. In part (c)(iii), many candidates did not realise that $\mathrm{P}(\text{at least } 21)=1-\mathrm{P}(\text{at most } 20)=0.083$. Even when this was realised, many candidates then thought that the required answer could be found by evaluating either 0.083^4 or $(0.083)^4\times(0.917)^2$. This gave scant regard to the fact that exactly 4 one-hour periods can be achieved from the given 6 one-hour periods in $\binom{6}{4}=15$ different ways. It was expected that $\mathrm{B}(6,\ 0.083)$ would be used here; this was often not the case.

Question 3

The hypotheses were either not stated or often incorrectly and unacceptably stated as "H_0: mean = 34.5 and H_1: mean $\neq 34.5$" **or** "H_0: $\bar{x} = 34.5$ and H_1: $\bar{x} \neq 34.5$" **or even** "H_0: = 34.5 and H_1: $\neq 34.5$". The forms of the two hypotheses which were acceptable and which were expected were either "H_0: **population** mean = 34.5 and H_1: **population** mean $\neq 34.5$" **or preferably** "H_0: $\mu = 34.5$ and H_1: $\mu \neq 34.5$". It should be noted that when hypotheses were not stated, statements such as "Accept H_0" were meaningless and so gained no credit. Also, a conclusion in context was required. Since the population standard deviation was given as 2.5, $z = \pm 1.96$ and **not** $t = \pm 2.009$ was required for the critical values. Some candidates seemed to think that they could use the *t*-distribution irrespective of what was given in the question.

Question 4

In part (a), many candidates were able to sketch the graph of f correctly. However, 'sketch the graph' does not give candidates licence to produce rough, freehand diagrams devoid of scales. In many cases, it was very difficult to give candidates credit since it was impossible to infer whether they intended to draw straight lines or curves. Consequently, some candidates lost some or all the marks for this part of the question. In part (b)(i), many candidates used $\int_0^2 \frac{2}{15} t \, \mathrm{d}t$, usually gaining the correct answer of $\frac{4}{15}$. However, as the graph contained straight lines, integration, although correct, was not the most efficient method. Although there were many correct solutions seen to part (b)(ii) by a variety of correct methods, there were still some candidates who treated $\mathrm{f}(t)$ as a discrete distribution and consequently thought incorrectly that $\mathrm{P}(2 < T < 4) = \mathrm{P}(T \leq 3) - \mathrm{P}(T \leq 2)$. This method gained no credit. Some candidates, presumably from experience on past papers, first derived the cumulative distribution function, $\mathrm{F}(t)$. Although this method gained the better candidates full marks, it was not the most efficient way of doing this part of the question. In part (c), the great majority of candidates considered the correct integrals and performed correct integration and were then able to evaluate these using correct limits to gain the required answer of $2\frac{2}{3}$. However, some candidates were unable to substitute the limits correctly in order to obtain this correct answer. Unfortunately, those candidates who thought $\mathrm{f}(t)$ to be discrete earlier in the question continued with this incorrect premise here.

Question 5

This was, somewhat surprisingly, the worst answered question on the paper. A **small** sample from a normal distribution with **unknown** variance required use of the *t*-distribution. Consequently those candidates, of whom there were far too many, who used $z = 2.5758$ instead of $t = 3.250$ inevitably lost marks here. The better candidates found $\bar{x} = 3.19$ and then correctly used $s^2 = \frac{1.849}{9} = 0.2054$ in constructing the 99% confidence interval by evaluating $3.19 \pm t_9 \times \frac{\sqrt{0.2054}}{\sqrt{10}}$ to give $(2.72, 3.66)$. Unfortunately, some candidates, who thought incorrectly that $s^2 = \frac{1.849}{10} = 0.1849$ and so used $3.19 \pm t_9 \times \frac{\sqrt{0.1849}}{\sqrt{10}}$ instead of

$3.19 \pm t_9 \times \frac{\sqrt{0.1849}}{\sqrt{9}}$, failed to find the correct confidence interval. In part (a)(ii), a statement that '3.5 is within the confidence interval' with a conclusion that 'the claim was therefore reasonable' was all that was required here. Part (b) was usually done well, except by those candidates who thought that 1% was equivalent to finding $\frac{1}{10} \times 200 = 20$.

Question 6

Most candidates apparently assumed that they had a **small** sample from a normal distribution with **unknown** variance and so used the t-distribution when conducting the required test. However, many such candidates either did not give an assumption or apparently did not know what assumption they had actually made. The assumption that "The sample is normally distributed" was the usual false statement and "It is normally distributed" or simply "Normally distributed" were also insufficient to gain the mark available. It was the **population** from which the sample was taken that had to be assumed to have a normal distribution. As in Question 3, hypotheses were sometimes either incorrectly stated or not stated at all. The correct values of $\bar{x} = 4.1$ and $s = 0.392$ were usually found and correct conclusions, in context, were usually seen.

Question 7

This question proved to be a good source of marks for most candidates. In part (a)(i), the great majority gained full marks. However, there was a minority who, having found correctly $\mathrm{Var}(Y) = 100$, failed to identify the standard deviation. Part (a)(ii) was done well by most candidates with $\mathrm{E}(C) = 255$ often seen. In part (b), the majority of candidates found the value for $\mathrm{Var}(X)$ and then went on to use $\mathrm{Var}(T) = 0.4^2 \times \mathrm{Var}(X)$ to gain the correct answer of 0.884. Those candidates who attempted to use $\mathrm{Var}(T) = \mathrm{E}(T^2) - [\mathrm{E}(T)]^2$, a valid if not the most efficient method, correctly found $\mathrm{E}(T) = 253.34$ but were then unable to show that $\mathrm{E}(T^2) = \mathrm{E}(0.16X^2 + 200X + 62500) = 64182.04$. As a result they often ended up with a negative answer for the variance.

Question 8

In part (a), $\mathrm{F}(0) = \frac{1}{k+1}$ was all that was required, but the use of $\mathrm{F}(0) - \mathrm{F}(-1) = \frac{1}{k+1}$ was not penalised. In part (b), most candidates realised that they needed to use the fact that $\mathrm{F}(q_1) = 0.25$ and so went on to determine the correct expression for q_1. Unfortunately many candidates must have thought that k was the lower quartile since they found an expression for k in terms of x. In part (c), candidates were expected to use the relationship $\mathrm{f}(x) = \frac{\mathrm{d}}{\mathrm{d}x}(\mathrm{F}(x))$ in order to gain the required given answer. Unfortunately, attempts at differentiating $\mathrm{F}(x) = \frac{x+1}{k+1}$ were, in the main, unconvincing with little or no attention paid to $\mathrm{F}(x) = 0$ and $\mathrm{F}(x) = 1$. In part (d)(i), most candidates gained full marks. Many candidates used integration methods in part (d)(ii) which, although correctly done, did waste a lot of time, especially as the answer for $\mathrm{E}(X) = 5$ could have been written down by simply considering the sketch drawn in part (d)(i).

Similarly, $\mathrm{Var}(X) = \frac{1}{12}(11 - -1)^2 = 12$ was a more efficient method of finding the variance.

There were many well-explained solutions to part (d)(iii) with most candidates realising that $\mathrm{P}(2 < X < 5)$ was required. Many candidates correctly stated that $\mathrm{P}(X < q_1) = 0.25$ but then assumed that $\mathrm{P}(X < \mathrm{E}(X)) = 0.5$ without any explanation. This statement was true here since, for a rectangular distribution, the median and the mean take the same value.

Mark Ranges and Award of Grades

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