General Certificate of Education

Statistics 6380

SS04 Statistics 4

Report on the Examination

2010 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

General

Candidates are increasingly using calculators instead of tables to produce cumulative probabilities, largely with success. The penalty for incorrect answers without sufficient evidence of method, however, can be severe. It was pleasing to see the majority of candidates attempt explanations to descriptive questions.

Question 1

The majority of candidates made good attempts at this question, using the normal distribution to find an approximate confidence interval for a Poisson mean based upon an estimated mean and a variance of 24, correctly interpreting the results and scoring full marks.

The most frequent errors in calculation were due to use of incorrect z values, with the one-tailed percentage point 2.3263 often seen instead of 2.5758. A few candidates incorrectly used Student's t-distribution on 23 degrees of freedom, and one or two used the estimated variance (24) divided by 7 (days per week) in calculation of the standard error.

Most candidates correctly stated that there was no evidence of a change in the mean number of births since 17 fell inside the confidence interval and hence the editor's claim of a (significant) increase was not justified. A few incorrectly asserted that there was evidence of an increase due to the large value for the upper confidence limit.

Question 2

Few candidates were able to precisely state an assumption for validity of the binomial model in part (a)(i), identifying a spelling error as the defined event assumed to occur independently with constant probability. "Random sample", "words independent" and "independent" were frequently seen in isolation and failed to score this mark.

Most candidates correctly identified a Poisson distribution having mean 0.7 as the appropriate approximation required in part (a)(ii), and used tables or calculators to correctly compute the probability of three or more errors. A common mistake, when using tables, was to compute the probability of two or more errors. The majority of candidates correctly identified a Poisson distribution with mean 274 as the model for grammatical errors in a 365-page manuscript in part (b)(i), either directly or, more often, implicitly through use of a N(274, 274) distribution in part (b)(ii). The most common incorrect answer was binominal (365, 0.75), subsequently followed through by an incorrect approximation in part (b)(ii). Many candidates failed to use correctly the continuity correction factor.

Question 3

Part (a) was well answered with many candidates scoring full marks. The sample mean and standard deviation were seldom calculated incorrectly, use of σ_n rarely seen, and consequently the correct test statistic (–0.435) was a very common outcome. Degrees of freedom were rarely quoted, the mark being implicitly earned through the usually correct critical value for Student's t, –1.833. Most candidates correctly concluded that Howard's claim was supported by the data.

Some candidates failed to state the null and alternate hypotheses, or failed to identify the lower one-tailed alternative. Sign errors were common, with negative test statistics compared to positive critical values. A few candidates chose to use the modulus of the test statistic compared to a positive critical value and this was considered inherently incorrect as a general approach to one-tailed tests and was consequently penalised. Critical values from standard normal tables, particularly –1.6449, were often erroneously used. A few candidates approached this question using two-tailed confidence intervals, and failed to gain full marks.

It was pleasing to see most candidates attempt full answers to the descriptive parts of this question.

A majority of candidates correctly identified that Jean suspected a type II error in part (b) although not all were able to correctly express this in context. Indeed, a considerable number appeared confused between type I and type II errors.

Very few candidates scored the second mark in part (c), giving a sufficiently detailed explanation as to why further testing was unnecessary (mentioning that the critical value is negative).

Most candidates correctly stated that reliability may be compromised by not taking a random sample in part (d) but, perhaps failing to read the question properly, lost the second mark by arguing this could be due to the selection of "underweight" bags.

Question 4

Overall this question was answered well by most candidates.

In part (a), the hypotheses were occasionally missing, the alternate hypothesis incorrectly stated (as two-tailed or $P > 0.2$), or inappropriate symbols used ($\mu = 0.2$ being common). Several candidates incorrectly used the sample proportion $p = 1/25 = 0.04$ to compute the p-value for the test, or used a normal approximation when an exact test was requested. The probability of exactly one attendee, $\mathrm{P}(1)$, as a p-value rather than the correct tail probability, $\mathrm{P}(\leq 1)$, was a common error. Most, but not all, candidates interpreted the test in context.

The majority of candidates correctly used the normal approximation to the binomial distribution to find the confidence interval in part (b). Incorrect forms for the standard error, usually a missing square root, were responsible for most incorrect solutions.

Those who attempted part (c) usually scored full marks. Incorrect solutions usually resulted from the use of $N = 1400$ rather than $N = 7090$ in calculating the expected number of attendees. A small number of candidates produced correct equivalent results by considering the proportion of members that could be accommodated by a room for 600 and comparing this to their confidence interval.

Question 5

Most candidates found this question straightforward and scored good marks.

In part (a)(i), an underlying binomial model was usually identified and a normal approximation used to test the hypothesis that 10% of guests arrived for breakfast before 7 am. Incorrect use of the sample proportion (26/234) to calculate the standard error in the test statistic (rather than 0.1) was a common error. Missing and incorrect continuity correction factors frequently led to an incorrect test statistic. Many candidates used calculators to compute the tail probability (p-value) rather than use the tabulated critical value (1.6449), but failed to compare their result for $1 - \phi(Z)$ to 0.05 or multiply by 2 (for a two-tailed test) when comparing with the significance level 0.10 (10%).

A number of candidates appeared confused when approximating one distribution by another and a normal approximation to a Poisson approximation of a binomial model was seen on several occasions. This was treated as a special case.

In part (a)(ii), candidates were invited to give reasons as to why the stated assumptions for validity might not hold. Simple comments that guests may not arrive independently with constant probability were insufficient to justify marks without qualification.

In part b(i), candidates correctly computed the sample mean and standard deviation with very few mistakes and usually gained full marks. Confidence intervals based on percentage points from the standard normal distribution (1.96), rather than Student's t, were the major reason for loss of marks. A very small number of candidates incorrectly used σ_n rather than s.

In part (b)(ii), several candidates lost the mark for failing to state the required assumption with sufficient clarity. "Normal" and "normal distribution" were frequently seen in isolation without qualification.

Surprisingly, very few candidates gained both marks for part (b)(iii), recognising that the confidence interval related to a population mean rather than individual data values.

Question 6

Most candidates who had sufficient time to complete this question gained good marks.

In most cases, part (a) led to full marks, with the incorrect addition of standard deviations rather than variances being the main cause for loss of marks.

Candidates found part (b)(i) a little more tricky. Many identified the need to calculate the probability of completing the cycling section in less than 16.5 minutes, or equivalently for total times for Fred and Ian to be less than 39 minutes based on means of 42 and 41.5 minutes respectively (conditional on having taken 22.5 minutes already). The first approach, not surprisingly, led to fewer mistakes. The most common cause of error resulted from the use of incorrect standard deviations. The use of 7.55 minutes for Ian from part (a) in particular was a common mistake with each approach.

Part (b)(ii) typically gained full marks. Where errors did occur this was largely due to identifying the wrong tail area required or, less often, incorrectly adding standard deviations rather than variances.

Mark Ranges and Award of Grades

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