Examiners' Report/
Principal Examiner Feedback

October 2016

Pearson Edexcel International A-Level
Statistics 2
(WST02)

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications are awarded by Pearson, the UK's largest awarding body. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information visit our qualifications websites at www.edexcel.com or www.btec.co.uk. Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

Pearson: helping people progress, everywhere

Pearson aspires to be the world's leading learning company. Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

October 2016
Publications Code WST02_01_1610_ER

Grade Boundaries

Grade boundaries for this, and all other papers, can be found on the website on this link:
http://www.edexcel.com/iwantto/Pages/grade-boundaries.aspx

General introduction

Questions 1 – 4 and 6 were a good source of marks for nearly all students whilst questions 5 and 7 discriminated at the top end. Students are encouraged to show their methods as much as possible even when calculations are carried out on calculators. There were a number of instances (2(d), 6(d)) where careless calculator work cost some students more than 1 mark as no indication of method was given. Students' work on hypothesis testing was generally pleasing whilst they were less confident applying the Uniform distribution in context.

Report on Individual Questions

Question 1

Most students made a good start to this paper with around half of them scoring at least 9 of the 10 available marks on this question. Students were confident applying their knowledge with the hypothesis test in part (a). The vast majority gave their hypotheses using correct notation, however, on some occasions, students incorrectly wrote the hypotheses in terms of λ or in words. The majority of students used the correct binomial distribution to calculate the required probability. The most common mistakes were to either use $\mathrm{P}(X > 4) = 1 - \mathrm{P}(X \leq 4)$ or to use $\mathrm{P}(X = 4) = \mathrm{P}(X \leq 4) - \mathrm{P}(X \leq 3)$. It was encouraging to see students giving a correct contextualised conclusion, though this was the most commonly dropped mark on the question.

In part (b) most students realised that a Poisson approximation was appropriate here and these went on to identify probabilities that would lead to a critical region. A frequent error was to incorrectly assume a 2-tailed test and hence give a 2-tailed critical region. It is still not uncommon to see a probability mistakenly given as a critical region. Students should make their reasoning clear as there were a number of cases where marks were lost when an incorrect critical region was stated without supporting probabilities.

Question 2
Students had the highest success rate with the second question on the paper almost half of all students scoring either 13 or 14 marks. Nearly all students provided a convincing argument to show that $k = -432$ in part (a). Most used the fact that F(12) = 1 to do so. In part (b) nearly all students correctly differentiated F(t) to find the probability density function, however some failed to include the range of t values for which it was valid. Another common mistake was to state that f(t) = 1 for $t > 12$.

In part (c) the majority of students simply wrote down the mode without any justification. The most common errors seen here were to state 12 as the mode or to write down f(8) as the mode. Part (d) had a higher success rate with most students able to set up F(m) = 0.5 and go on to solve the resulting quadratic. Use of a calculator was common, but students are reminded that their method must still be made clear.

Again, part (e) was well answered with most working out F(9) to calculate the required probability. On some occasions students incorrectly believed the distribution to be discrete and calculated $\mathrm{P}(T < 9)$ as $\mathrm{P}(T \le 8)$. The most discriminating part of the question came in part (f) where many students failed to appreciate that a conditional probability was required. Those who made some progress attempted $\mathrm{P}(9 < T < 11)$, but often went no further. Another commonly seen mistake was to try and use $\mathrm{P}(T < 11 \mid T > 9) = \dfrac{\mathrm{P}(T < 11)}{\mathrm{P}(T > 9)}$

Question 3
Question 3 was more demanding with only 20% of students achieving full marks. In part (a), the majority of students identified the correct binomial distribution including n and p. The inequality signs in part (b) caused more difficulty for some students as many incorrect attempts were made at $\mathrm{P}(4 \le X < 9)$. These included $\mathrm{P}(X \le 8) - \mathrm{P}(X \le 4)$ and
$\mathrm{P}(X \le 9) - \mathrm{P}(X \le 3)$.

In part (c) some students solved an inequality whilst others made a table of possible total points scored, but almost all students worked out that more than 6 correct questions were required to achieve a total score greater than 0. There were fewer problems with inequality signs here as most went on to calculate $1 - \mathrm{P}(Y \le 6)$.

Part (d) was the most discriminating part of the question. Most students simply worked out the variance of X, but went no further. In order to calculate the variance of the total points scored in a random sample of 20 students, it was necessary to come up with an expression, in terms of X, for the total number of points scored. Those who did so generally had success in multiplying Var(X) by 10^2 to obtain the correct answer.

Question 4
Most students found question 4 accessible and nearly 40% earned full marks here. In part (a), the straightforward method involving areas was not always used with some opting for integration. It was, however, pleasing to see a small number of students using a trapezium area approach. Those integrating did not always make the connection between m and k so often got stuck. For the most part, m was correctly found.

In part (b) the approach to finding the mean is generally well known with many correct answers seen. Some slips included multiplying only one of the integrals by t before integrating but not the other. Another common error seen was a correct integration for the first part of the function then added to the mid-value from the rectangle. Correct methods which did not involve integration were very rarely seen with most of these attempts scoring no marks.

In part (c), the lower quartile was successfully found by many. An integration approach was most popular instead of using area on the diagram. Those using integration for the upper quartile often made the common mistake of not including the first part of the function. The simple approach in this question of finding Q_3 by taking an area of 0.25 from right limit of the rectangle was rarely seen. A few students successfully found the cumulative distribution function first, but did not always go on to use it correctly. Quite a few, after finding both quartiles correctly, gave the interquartile range as an interval rather than the difference between the quartiles.

Question 5
Question 5 was found to be the most challenging on the entire paper with ¼ of students making very little or no progress. Setting up the correct distribution in part (a) was challenging for many students. Some were able to gain 1 mark for identifying the Uniform distribution, but many went on to give the incorrect lower boundary for the longer piece of string. Those who did write [0, 40] were able to earn follow through marks in the latter parts of the question. Some believed that a Poisson distribution was appropriate and often struggled to pick up any marks on this question.

Of those students who understood that the required probability was $P(27.5 < L < 28.5)$ in part (b), most went on to use a correct method to find this probability. Many gave an answer of 0 attempting to display their knowledge of a continuous distribution but neglected the phrase ‘to the nearest cm’. Others incorrectly attempted $P(27.5 < L < 28.4)$.

Even those without any progress in parts (a) and (b) were able to obtain a correct expression for L in part (c). Some students forgot to divide by 4 before calculating the area and obtained $L < 8$. Part (d) was a challenge to some students as there were often slips in writing down the initial expression for the difference in the two areas. The resulting expansion was not always carried out accurately. Some still scored the second method mark by applying the uniform distribution correctly. Overall, parts (c) and (d) were two of the most discriminating parts of the entire paper.

Question 6

There were a number of very good responses seen to question 6 with 20% of students scoring full marks. There were many good answers to parts (a) and (b) with the majority recognising the appropriate Poisson distribution in each part and stating each correct mean in terms of λ. A few careless misreads led to students calculating $P(X = 2)$ or $P(X \leq 2)$ in part (a).

Part (c)(i) was often not answered at all whilst others misunderstood the demand altogether. Common non-scoring responses were to state 'λ was large', or that 'mean = variance', or to simply write the approximate normal distribution. The minority who were correct often stated that $10\lambda > 10$.

A large number of fully correct responses to (c)(ii) were seen with nearly all identifying the correct normal approximation. These often went on to apply the correct continuity correction and standardise correctly. There were some who made errors in forming a correct statement for z either using an inconsistent sign for their z-value or, in a few cases, equating to a probability. Those forming the correct standardised equation generally recognised and solved the resulting quadratic equation to obtain the correct answer. The single correct root was generally isolated. A small number either did not recognise it as quadratic or were ill-equipped to solve it, trying to rearrange unsuccessfully. Having reached a quadratic equation, the solution method is often not shown preventing the award of a final method mark if the correct answer was not reached. As in question 2(d), students are perhaps relying on quadratic solution functions on calculators but are reminded that they should make their methods clear.

Question 7

There was a mixed response to question 7 with 15% of students making no progress at all and 30% scoring full marks. Overall this proved to be the second most demanding question on the paper. In part (a), most students were able to identify all of the possible samples. A significant number then went on to attempt the sampling distribution of the mean number of ice cream scoops rather than the total. Most attempts at the required probabilities for $S = 2$ and $S = 6$ were correct. There were more slips seen with the other probabilities most notably $P(S = 4)$. It was pleasing to see well-presented sampling distributions.

Many found part (b) challenging and there were many blank responses seen here. For those who persevered and understood what was required, most were able to set up $1-\left(\frac{5}{8}\right)^n > 0.99$ and generally had success in solving this inequality. Occasional slips with the inequality sign led to a common incorrect answer of $n = 9$. Here students are again reminded that showing their working is essential to ensure that method marks can be earned.

Pearson Education Limited. Registered company number 872828
with its registered office at 80 Strand, London WC2R 0RL