Examiner's Report
Principal Examiner Feedback

October 2018

Pearson Edexcel International A Level
In Statistics S2 (WST02/01)

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 2018
Publications Code: WST02_01_1810_ER
All the material in this publication is copyright
© Pearson Education Ltd 2018

General

This paper proved to be accessible to all students though questions 2(c)(ii), 3(a), 5(c), 6(a) and 7(e) were more discriminating. Students should make their methods clear, even when using a calculator to solve equations. They should also ensure that answers are not rounded too early as it may affect the accuracy of latter parts of questions. In general, students struggle with questions involving conditional probabilities.

Question 1

The opening question of the paper was accessible to students of all abilities and over 40% achieved full marks. In part (a) nearly all students identified Po(6) and the majority were able to find the $P(X=1)$ to an appropriate degree of accuracy. On some occasions a less accurate answer of 0.015 was given. Students should be well aware that answers should be given to at least 3 significant figure accuracy unless otherwise stated.

Part (b) was also generally well attempted. Most students knew to state their hypotheses using correct notation λ but some students used p or did not label their hypotheses correctly whilst others thought that 14 was the value to use. The vast majority of students opted for comparing the relevant probability with 0.05 rather than finding the critical region. The usual mistakes were finding $P(X=14)$, $1 - P(X \leq 14)$ or $1 - P(X \leq 6)$. Most conclusions were given in context.

Question 2

In this question students showed their confidence with the binomial distribution. The conditional probability required in part (c)(ii) discriminated the most able.

Part (a) was nearly always answered correctly with only a handful of errors seen. These included finding $P(X \leq 3)$ or attempting $1 - P(X \leq 2)$ from B(12, 0.2).

In (b)(i) most were able to clearly show that the probability was 0.4 and many used a tree diagram to aid their calculation. A number of students left the probabilities as percentages in their expressions. Common incorrect answers to (b)(ii) included ‘uniform’ or ‘continuous’. Nearly all who identified the binomial distribution went on to define it fully.

Again, (c)(i) was straightforward for most. It was (c)(ii) that caused the most difficulty as correctly working out the numerator of the conditional probability proved a challenge for many. Often students found $P(X \leq 6)$ and then realised that their denominator was too small and gave up. Some incorrectly interpreted $P(3 \leq X \leq 6)$ as $P(X \leq 6) - P(X \leq 3)$. There were also slips in the denominator with $1 - P(X \leq 3)$ sometimes seen.

There was greater success in part (d) as most found the correct Normal approximation to the binomial. Though the use of the continuity correction is generally well known, some used 75.5 instead of 74.5. Nearly all students showed their standardising and gained method marks even following on from other errors.

Question 3

This question was significantly more challenging than the opening two with only 10% of students managing full marks here. In part (a) sketches of f(x) were generally of a poor standard. Many neglected the range and drew the entire quadratic. Others tried to plot point by point, often resulting in a straight line instead of a curve. Those offering a reason why f(x) could not by a probability density function often had flawed logic. Many said that probabilities cannot be greater than 1 demonstrating a misunderstanding as to what a p.d.f. shows. Another common incorrect comment focused unnecessarily on the fact that the graph was continuous.

Most students were able to use differentiation to find the mode in part (b). A few did not use differentiation but calculated g(1), g(2) and g(3) and used these values to take a guess of the mode. Sometimes students had difficulty expanding g(y) or differentiating correctly.

Students were most confident with part (c) of this question and most completed this accurately displaying knowledge that the integral of the p.d.f. must equate to 1.

In part (d), many were able to find the median with a correct integration and accurately solving the resulting polynomial to obtain the correct solution (and reject the solution out of range). Some students incorrectly used $\int g(y)dy=0.5$ with no limits nor any attempt to include and evaluate an arbitrary constant. Some in fact thought that g(0.5) was the median.

Most students were able to compare their mode and their median in part (e). Some tried to evaluate the mean, although this was not required, and use this in their comparison. Those who had incorrectly found a median or mode outside the domain were not able to gain credit for their comparison. Notably very few students judged that the 'mean ≈ median' and concluded the distribution is 'almost symmetrical' or had 'no skew'. Most opted to state that the distribution had a negative skew.

Question 4

There was mixed performance on this question with nearly ¼ of students making no progress yet ¼ went on to score full marks. Some attempted only parts (a) and (d) showing that they were not fully prepared for all aspects of the specification. Most showed a full expression of the required probability in part (a)

Part (b) was generally fully correct with students demonstrating full understanding of the concept of a sampling distribution or it was not attempted at all. A few students could not work out the values that D could take on but were able to recognise the associated probabilities.

The mode was fairly well attempted in part (c) with many understanding that it was the value of the random variable with highest probability. Many, however, thought it was the event that had the most ways of occurring and gave both 1 and –1 as the mode.

Most students have a solid grasp of the process of finding the critical region and were able to use the correct distribution in part (d). However, a significant minority assumed this was a two-tailed test. Students should be reminded to write down relevant probabilities to support their critical regions as this will help ensure method marks are scored. A still common mistake is to write the critical region as a probability. Others did not consider a 10% level of significance, often opting for 5% or 1%.

Many students were able to benefit from the follow through mark in part (e) when they considered the significance level of their one-tailed test.

Question 5

There were many strong performances on this question as the relationship between the c.d.f. and the p.d.f of the continuous random variable is well known. There were, however, a significant number of students who made little or no progress by attempting to integrate F(x). Those using correct mathematical notation generally made their method clear and worked through the steps logically to prove the given result. Many started by using the fact that F(5) = 1 although this was not required in part (a). Others tried to overcomplicate things by coming up with an expression for the variance of X. Still a minor misconception is that $E(X^2) = [E(X)]^2$.

Students were generally able to gain full credit in part (b) even following errors from part (a). Most were able to identify the second equation using F(5) = 1 and when they did so solving the simultaneous equations was straightforward. Method marks can only be awarded if working is shown, so if students did obtain the wrong equations and used their calculator function, they were unable to score the method mark here.

Part (c) had the lowest success rate. Two errors were common. The first was the candidate who did not recognise that F(7) = F(5) = 1 and the second was the candidate who treated X as a discrete random variable and incorrectly wrote $P(3 \leq X \leq 7) = F(5) - F(2)$.

Question 6

This question on the continuous uniform distribution discriminated the most able students with parts (a) and (b)(ii) being the most demanding. It was uncommon to see a correct explanation in part (a) as to why the random variable was distributed over the interval [–2, 2]. A common error was to say $a^2 = 4$ so $a = \pm 2$. Others mentioned that a square has 4 sides, but took no notice of the error in the measurement of one side length.

Part (b)(i) was well answered as most students used the complete correct formula for standard deviation. A few, however, forgot to take the square root of the variance. Part (b)(ii) was not as well attempted as most could not interpret the probability that was required. Of those who had a valid attempt, this often lacked accuracy due to using a rounded value from the previous part.

In (c), many students did not recognise that the Poisson distribution was required here. Most made some progress by finding $P(X \geq 1)$ from the uniform distribution. Some students did go on to write B(100, 0.025) but this was often followed with a Normal approximation.

Question 7

The opening parts of the final question of the paper were well attempted, but part (e) was a challenge for most.

It was very encouraging to see students use correct notation in part (a) as many wrote write $P(X \geq n) < 0.05$ and then found the answer of 4. This part was well answered.

There were a few more slips in part (b) as some neglected the fact that the rate was $0.5m$. This was generally well answered as many students wrote down $e^{-0.5m} < 0.05$ and then obtained the correct answer of 6. Some students used $1m$ or $1.5m$ for λ.

Part (c) was significantly more demanding as many students believed that they could work out the required probability directly using Po(1.5) rather than using $(1 - P(X \geq 1))^3$.

Many wrote down Po(4) in part (d) but did not use this value to answer the question.

Finally, part (e) was only tackled by the most able students. Only a minority gained any marks and generally this was for a correct denominator. Most students could not extract the conditional probability ratio expression from the information given. Many did not realise that in order for both events to be true the numerator required a sum of 4 sightings. Thus $P(A \geq 2)$ was often used as the numerator, but this led to a probability greater than 1.

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