Examiner's Report
Principal Examiner Feedback

October 2019

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 2019
Publications Code: WST02_01_1910_ER

General

This paper was accessible to all students but there were several places where students struggled to translate the context into correct statistical processes/calculations. Whilst parts of all questions were accessible, only the most able students made full progress with the most demanding questions on the paper, questions 5, 6 and 7.

Question 1

(a) Almost all students identified that this was a Binomial distribution of B $(4, p)$, however, only a small minority realised that it was easier to find P $(X = 3)$ + P $(X = 4)$. The majority of students found $1 - \text{P}(X \leqslant 2)$, with most losing a mark as the required lines of working were not shown.

(b) This was well attempted with most students attaining full marks; however, the common errors were not squaring the standard deviation or not rejecting the probability that was less than 0.5

(c) As is often the case, the conditional probability was not identified and the majority of students simply calculated P $(X = 3)$, resulting in no marks. When conditional probability was identified, in about one quarter of students, these students gained full marks.

Question 2

(a) This part was generally answered very well and was accessible to a very large majority of students. It was encouraging to see many write down their method clearly to show a correct answer of 8.25

(b) This was generally done well. Many students realised that $\text{P}(5 \leqslant X \leqslant 14)$ was equivalent to $\text{P}(5 \leqslant X \leqslant 12)$. Common incorrect answers were $\frac{13-5}{12}\left(=\frac{8}{15}\right)$ or $\frac{14-5}{12}\left(=\frac{9}{15}\right)$.

(c) This was answered well by many students. The most common error in finding E(Y) was forgetting to add the 3 to 4.5 to obtain E(Y) thus losing the first mark. It was encouraging to see many students obtain 18.75 or 4.6875

Common errors in finding the equation using Var (Y) were to write down $4^2\text{Var}(X)$ rather than $4\text{Var}(X)$ or using a $\frac{1}{4}\text{Var}(X)$. Some students involved complicated quadratics and substitution without realising that the quadratic can be simply square rooted to obtain the correct second linear equation. Even those that managed to set up two correct linear equations continued with substitution rather than simply adding the two equations.

Question 3

(a) This was a "show that" question which was very well answered. The majority of students showed their working clearly and gained full marks.

(b) Although there were a substantial number of correct solutions, there were quite a few errors. Mostly by students thinking they needed to use a different distribution from Poisson - usually normal which led to statistical gymnastics of a strange kind.

(c) The hypotheses were generally correct. The most frequent error was writing H_0 as μ >50.4 The normal approximation was recognised by the majority and most used a continuity correction. A minority used no continuity correction or 37.5 rather than 38.5

The conclusions were well written with the majority including the relevant context.

Question 4

This was a long and a challenging question for some students but those students who laid out their work clearly and communicated their reasoning methodically were most successful.

(a) A large majority of students were familiar with the theory but were also able to write down accurately the required definite integration with only an occasional error seen in the arithmetic. There were, however, a very small number of students whose entire working for the first integral, for example, consisted only of: $\int_3^x \frac{1}{15}\left(3x^4 - x^5\right)dx = 1.59$ ignoring the instruction to "use algebraic integration" and this resulted in a loss of marks. A small number of students made the predictable error of integrating f(t) for both parts although it was surprisingly common to see students perform only one of the two integrals. Although a few errors were made in the integration most resulted from inaccuracies when it came to substituting in the limits. Many students did not show the substitution clearly and in these cases arithmetical errors were often made. Only a minority of students correctly evaluated both integrals and stopped before adding them.

(b) It is reassuring that many perfect solutions were seen for this routine type of question. There were the usual slips with the algebra and the arithmetic. The most common problem was the part of the distribution dealing with3 $\leqslant x \leqslant 5$ where many students simply worked out $\int_3^x \frac{3}{10}(t-3)\,dt = \frac{3}{20}t^2 - \frac{9}{10}t + \frac{27}{20}$

Students who obtained the correct answer used a mixture of methods. Some used the approach $\int_3^x \frac{3}{10}(t-3)\,dt$ + F(3) where F(3) = $\frac{2}{5}$, while others chose indefinite integration: $\int \frac{3}{10}(t-3)\,dt = \frac{3}{20}t^2 - \frac{9}{10}t + c$ and F(5) = 1

(c) This part was often completed well but there was some confusion between discrete and continuous distributions. The most common errors were using P(2 < X < 4) = P(2 < X < 3) or P(3 < X < 4) or F(3) – F(2) or F(4) – F(3). A small number of students simply evaluated f(4) – f(2) using the probability density function not the cumulative distribution function.

(d) Many students obtained the correct answer of 4.63 with 1.37 clearly crossed out. The most common error was to amalgamate their two functions from part (b) and then equate it with 0.2 or 0.8. Some students used the incorrect function, i.e. their function from (b) for $1 \leq x < 3$ and then equated this with 0.8. Those students who used calculators to solve the quadratic equation usually provided the required amount of supporting detail.

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

(a) As this part referred to a uniform distribution it was usually correctly calculated in a straightforward way involving a simple expression and/or an area on a graph of the density function. However only a few effectively used a graph and others made the problem a lot more complex than it need have been, perhaps by first writing out the cdf. A number of students found the probability that the bolts couldn't be used rather than the probability that the bolts could be used.

(b) Most students correctly used a $B(10, p)$ distribution here. It was possible to use either $B(10, 0.95)$ or $B(10,0.05)$ with the probabilities $P(X \geqslant 9)$ and $P(X \leqslant 1)$ respectively. This resulted in an evident confusion in the minds of many students as they frequently wrongly calculated $P(X \geqslant 9)$ with the distribution $B(10, 0.05)$

The simple use of the binomial function appeared to be somewhat challenging and many students wanted to use tables instead.

(c) Students found this part of the question difficult.

As the question was framed around 'bolts that can be used' most students began by writing $X \sim B(120, 0.95)$

As they were asked to use an approximation, they continued by calculating a mean '*np*'. As this was 'large' (114) they then proceeded to use a normal approximation without considering that the conditions for approximating to a normal were not valid. (i.e. p was not 'close to' $\frac{1}{2}$).

It was essential for this part that the student changed to count the bolts which could not be used. For some who correctly wrote $Y \sim B(120, 0.05)$ and approximated to Po(6) they still had difficulty converting $P(X > 117)$ to $P(Y \leqslant 2)$.It may have been easier for some if they had written something like $P(X = 118, 119, 120) = P(Y = 2, 1, 0)$

Question 7

(a) The majority of students correctly found $\frac{\mathrm{dF}(x)}{\mathrm{d}x}$. Many then correctly differentiated again to find $\mathrm{f}'(x)$ and went on to solve $\mathrm{f}'(x) = 0$

However, there were some who incorrectly solved $\mathrm{f}(x) = 0$ instead. Others realising that they needed a maximum evaluated $\mathrm{f}(x)$ at whole number points only giving them a mode = 4
(b) This part was well done by many students, even when there had been problems earlier in the question. Most understood what they were trying to do, and often had a correct version of the function to hand. However, some failed to provide an acceptable conclusion: "so the median is between the two numbers" or "$\mathrm{F}(3.95) < m < \mathrm{F}(4.05)$" are examples which did not gain the mark. The most common error was to amalgamate their two functions from part (b). Others used the 'wrong' function, i.e. their function from (b) for $1 \leqslant x \leqslant 3$. Those students who used calculators to solve the cubic equation did not provide the required amount of supporting detail.
(c) It appears the words 'more than' in the question caused many students to incorrectly state $H_1: p > 0.25$. If they followed this through and found $P(X \geq 3) = 0.9679$ they normally compared this to 0.05 and did not seem to realise that there was an issue with finding a probability close to 1 when usually a test would have given them a small probability. Most students who had the correct $H_1: p < 0.25$ and correctly calculated $P(X \leqslant 3) = 0.0962$ were then able to follow with a good conclusion in context.

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