Examiners' Report
Principal Examiner Feedback

January 2021

Pearson Edexcel International A Level

In Statistics 2 (WST02)

Paper: 01 Statistics S2

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

January 2021
Publications Code WST02_01_2101_ER

General

Candidates were generally well prepared for this paper with strong statistical knowledge often displayed. In order to earn full marks, candidates must show all stages of their working, particularly when the answer is given as was the case in 3(c)(i), 4(a)(b)(c) and 6(d). When working with their calculators, it is vital that candidates make clear which distribution they are using and what probability they are attempting. This will ensure access to method marks when there are slips in accuracy. Questions 6(b) and 6(c) were often left blank showing that candidates need to enhance their work on understanding statistical vocabulary associated with sampling.

Comments on the individual questions

Question 1

Question 1 provided a solid start to the paper for a majority of the candidates.

Part (a) was answered well. A common error was to write down 'binomial' on its own which was not sufficient.

In part (b) many candidates earned the mark by writing answers in context. Answers which did not contain sufficient context or those restating facts given in the question (such as 'there are only 2 outcomes') did not score here.

Most answered (c)(i) correctly but a common error was writing their expression as ${}^{30}C_6\left(0.05\right)^{24}\left(0.95\right)^6$. In (c)(ii) it was possible to use either B(30, 0.05) or B(30, 0.95) with the probabilities P($X \geq 3$) and P($X \leq 27$) respectively. This resulted in confusion for some as a number of candidates attempted P($X \geq 3$) with the distribution B(30, 0.95). Some candidates incorrectly used P($X \geq 3$) = 1 – P($X \leq 3$).

Most candidates made it clear that they were using a Poisson approximation in part (d) although a few used a Normal approximation and some failed to make it clear which distribution they were using. The method mark was usually obtained with P($A \leq n - 1$) < 0.2 although the answer was sometimes given as n = 4, presumably because they had misread 0.2 as 0.02.

The hypotheses were usually correctly stated in part (e) with only a few candidates forgetting to use p and almost all getting the inequality in H_1 correct. Many candidates went on to calculate the correct probability but a common error was to find 1 – P($X \leq 4$) or P(X = 4). Another error made by the candidates was to find P($X \geq 27$) rather than P($X \leq 27$). Those who obtained the probability of 0.0341 correctly compared it with 0.05 and gave a correct conclusion. The contextual statement was usually present and with sufficient clarity and detail to gain the final mark. Occasionally a 2-tailed test was attempted but these often did not compare their probability with 0.025.

Question 2

Question 2 provided a good source of marks for candidates at all levels.

A large proportion of candidates scored both marks in part (a). The rare incorrect methods included F(3.5) and 1 – F(2.5). A less serious error involved unsuitable rounding to 0.029, or extremely rarely, 0.03. A few candidates obtained the correct answer by the inefficient method of first differentiating to find f(x), and then integrating with limits to obtain the required probability.

Full marks were obtained in part (b) by a majority of candidates who achieved the attention to detail required for accurate use of the formula for a Binomial probability. The most serious difficulty concerned the phrase: "more than one". It was not uncommon to see the incorrect method 1 – P(X = 0). Some, using their calculator, did not make their method clear and lost all three marks.

A substantial number of candidates demonstrated a sound understanding of the underlying principles in part (c), which led to many clear, accurate and concise solutions scoring full marks.

The most significant difficulty in part (c) was the failure to differentiate the cumulative distribution function in order to obtain the probability density function. As a result, the incorrect method $\int w^2 \mathrm{F}(w) dw$ was often seen.

A few candidates failed to follow the instruction in part (c) of the question: "use algebraic integration ..." and were unable to access the second method mark in this part. Sometimes incorrect integration was seen followed by a correct answer, likely obtained by a calculator.

Question 3

Part (c) of this question discriminated the most able candidates, but all were able to access marks here and make some good progress.

Almost all candidates were successful in part (a) although some sophisticated methods such as $1 - (P(X \leq 4) - P(X \leq 3))$and $1 - P(X \leq 4) + P(X \leq 3)$ were seen.
Part (b) was generally well done, with many responses achieving full marks. Some candidates were able to obtain partial credit in (b) by using an incorrect answer from (a). However, this was only possible if full working were shown, in particular a detailed expression for the Binomial probability. Some lack of accuracy resulted by those candidates who prematurely rounded their answer to part (a).

The overall response to (c)(i) was satisfactory. All five marks were indeed scored by many candidates who displayed attention to detail throughout. Perhaps the most serious problem occurred at the outset: the failure to identify the correct Normal distribution. A relatively small number of scripts included a sign error in the Standardisation equation. It is interesting to note that the correct continuity correction was used by the vast majority of candidates.

Many different versions of algebraic manipulation were seen in the final step from the standardisation equation to the printed answer. Some excellent solutions clearly and confidently distinguished between arithmetical operations (specifically multiplication and division) that occur outside or inside a square root sign.

Candidates were required to solve a quadratic equation in part (c)(ii). This may sound like a routine operation, but an extraordinary range of responses were in evidence. One issue was that this is a 'disguised' quadratic. It appears that a few candidates failed to recognise this and made no attempt to solve the equation. Some candidates simply used the quadratic formula to solve for $\sqrt{n}$, but then failed to square this answer to obtain the final answer.

A substantial number of candidates earned full marks in part (d) stating their hypotheses clearly with correct notation. Only a relatively small number of candidates failed to use the correct Poisson distribution or failed to identify correctly the event "15 or more fleas". The latter was the most serious error leading to a loss of marks. Almost all candidates provided a final conclusion with appropriate context.

Question 4

On the whole, candidates dealt well with the demands of the algebraic integration involved in this question.

Part (a) was generally well done with a clear layout and sufficient working shown to establish the proof. Candidates almost always multiplied out the brackets before integrating. Only a few attempted to integrate 'by inspection'. Those that did often omitted the minus sign outside the bracket. It was encouraging most of the candidates integrated correctly. However, some integrated with respect to both the variable a and x, for example, writing down

$$\int_0^a ka^2 - 2kax + kx^2 dx = \left[\frac{ka^3}{3} - \frac{2ka^2x}{2} + \frac{kx^3}{3}\right]_0^a.$$

Part (b) was also well answered by the majority of the candidates. As above, the candidates showed a clear layout and sufficient working shown to establish the proof. Many candidates used the given information that $\mathrm{E}(X) = 1.5$ and correctly multiplied $x\mathrm{f}(x)$ prior to integrating, but often an elementary algebraic error prevented candidates from reaching a correct equation. There were again problems regarding the variable of integration. Some failed to show the substitution of the limits leaving out an important step to 'show' the given answer.

There was a somewhat mixed response to part (c). Sometimes it was very well done with candidates clearly well versed in solving this type of problem. The majority could work out the correct expression for $\mathrm{F}(x)$. Successful candidates went on to work out F(1.15) and F(1.25) were able to explain why this showed that the median was 1.2 to 1 decimal place. Some used their calculator to solve $\mathrm{F}(m) = 0.5$ and were also able to explain correctly why $m = 1.2377\ldots$ meant that the median rounded to 1.2. However, many who used their calculator did not consider writing down a more accurate solution than 1.2 losing the final mark. Quite a few merely asserted that $\mathrm{F}(x) = 0.5$ implied that the median was 1.2 There were some elegant solutions seen using $\mathrm{F}(x) = 1 - \frac{(6-x)^3}{216}$ leading to an exact value for the median of $6 - 3\sqrt[3]{4}$. Some candidates simply evaluated$\mathrm{F}(1.2)$to give an answer just below 0.5. But this is not enough to demonstrate that the median is in fact 1.2.

Question 5

This proved to be one of the most demanding questions on the paper with the most able candidates persevering and making good progress.

Most candidates not only recognised that a Uniform/Rectangular distribution applies in part (a) but were also able to use it effectively. Other candidates decided that a Poisson distribution was appropriate. There were yet others who used a distribution with incorrect parameters, the most common being U[1.5, 3].

Part (b) was generally well done with sufficient working shown to establish the proof. Many wrote down that $CB = 3 - W$ and then used Pythagoras' theorem correctly to lead to the given result.

There was a mixed response to part (c). Some candidates mixed up the variable X and W. A common error was to work out $\mathrm{E}(W^2) = \frac{(3-0)^2}{12} = \frac{9}{12} = \frac{3}{4}$ and then using this value in the equation $\mathrm{E}(W^2) = \mathrm{Var}(W) + [\mathrm{E}(W)]^2$ leading to an incorrect answer. Some candidates wrongly used $\mathrm{E}(W^2) = [\mathrm{E}(W)]^2$ Some candidates assumed that X^2 itself had a uniform distribution. A small minority of candidates used the alternative method of integration and had an equal success rate.

(d) Generally, the most able candidates could gain full marks. Many did get to the stage $\mathrm{P}(X^2 > 5) = \mathrm{P}\left(2W^2 - 6W + 4 > 0\right)$ and then worked out that $W = 1$ and $W = 2$ thus gaining 2 marks. Some obtained $1 < W < 2$ as the solution to the inequality or $W > 2$ and $W > 1$. Many were not able to earn the final 2 marks as they did not know that they had to add the probabilities of $\mathrm{P}(W > 2)$ and $\mathrm{P}(W < 1)$ to obtain the correct answer. Some candidates multiplied the probabilities rather than adding them. Some blank scripts were seen for part (d). Other candidates were unable to make any progress beyond $\mathrm{P}(X^2 > 5) = 1 - \mathrm{P}(X^2 \leq 5)$.

Question 6

Part (d) was one of the more demanding parts of the paper, but there were a significant number of candidates who did not make any attempt at parts (b) and (c) despite these being standard definitions.

Part (a) was the most successful of the first three parts, followed by (c). A lack of detail meant the mark in part (b) was often not scored. A sampling frame must include **all** sampling units, in this case youth club members. Many tried to restate textbook definitions rather than to try and engage with the context of this question.

There were three unknowns in part (d). The two pieces of information printed in the question lead to the equations $p^2 = \frac{25}{64}$ and $2qr = \frac{1}{16}$

A third equation was therefore required. It was expected that candidates would apply the principle that the sum of probabilities is equal to 1 to the sentence in the question: "Each ball is numbered 20, 50 or 70" and realise therefore that $p + q + r = 1$. This approach lead to two equations in q and r: $q + r = \frac{3}{8}$ and $qr = \frac{1}{32}$. Candidates obtained and then solved a quadratic equation of the form $q^2 - \frac{3}{8}q + \frac{1}{32} = 0$ or $32\mathrm{q}^2 - 12\mathrm{q} + 1 = 0$.

A rather more complicated solution saw candidates apply the principle of total probability to the 'sampling distribution of M'. This was a fairly common approach. This is indeed a valid, alternative method that led to the correct final answer. However, a substantial amount of additional work was required using this method. Instead of one of the two equations above in q and r, one possible outcome is the complicated equation printed in the mark scheme, featuring unknowns in q,q^2,r,andr^2. Solving simultaneous equations, with this as one of the pair was significantly more challenging as it leads to a quartic equation, typically

$$q^4 + \frac{5}{4}q^3 - \frac{35}{64}q^2 + \frac{5}{128}q + \frac{1}{1024} = 0$$

Most used the given answer to verify that $q = ¼$ and then went on to complete the solution.

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