Examiners' Report
Principal Examiner Feedback

January 2023

Pearson Edexcel International Advanced Level
In Statistics S2 (WST02) Paper 01

<u>Report on Individual Questions</u>

Question 1

Most candidates made a very strong start to the paper with this question. Incidentally the most common problem occurred in part (a) with the absence of the parameter 4.

A large majority of candidates had the right general idea about part (b), although some omitted any reference to the specific context.

Fully correct solutions to part (c) were often seen. The most common error in part (c)(i) was perhaps the misconception that $\mathrm{P}(X < 4) = 1 - \mathrm{P}(X \leq 3)$.

A majority of candidates realised that the cube of (c)(i) would provide the answer to (c)(ii). One serious (but very rare) misconception was to multiply the answer to (c)(i) by three. More common were attempts that featured the Normal distribution N(12, 12) and the probability $\mathrm{P}(X > 4)$, sometimes even with a 'continuity correction'.

Part (d) was well done with the majority of candidates gaining full marks. There were very few errors with the hypotheses, where errors were made it was when hypotheses were written in sentences or using 7 as opposed to 4. Most candidates were able to identify the correct p-value of 0.1107 (and very few took the CR approach) and their conclusion was contextualised adequately in most cases. One common mistake was to find the probability $\mathrm{P}(X > 7) = 1 - \mathrm{P}(X \leq 7)$ with a small minority of candidates attempting $\mathrm{P}(X = 7)$.

Question 2

This question on sampling distributions was well attempted by the majority of candidates. Parts (a), (b) and (c) were generally fully correct. Most candidates gave all 8 possible permutations. There were a couple of instances where the word 'list' was ignored from the question and the answer was presented as a tree diagram, for example.

More commonly, trouble only started in part (d). Even then, there was only one major problem, which was relatively rare: the omission of the Binomial coefficient $\binom{3}{1} = 3$.

Part (e) proved the most challenging part of the question as some attempted the sampling distribution of the median instead of the mode. In this part most candidates displayed their answers in a table and the probabilities were usually correct although a few gave both probabilities as ½. Failure to add their values of *c* and *d* to the given probabilities was seen, but the most usual reason for losing the A mark was having wrong values for *c* and *d* and failing to notice their two probabilities did not sum to 1.

Question 3

Many candidates scored full marks for part (a). This was not, however, universal. The inequalities in part (a) were not well handled by all the candidates. The common error in(a)(i) was $\mathrm{P}(X \geq 4) = 1 - \mathrm{P}(X \leq 4)$. A wide variety of errors were seen in (a)(ii): incorrect attempts at $\mathrm{P}(1 < X < 5) = \mathrm{P}(1 \leq X \leq 4)$ included all versions consisting of a value either side of 4 and/or a value either side of 1.

Other candidates complicated matters with incorrect statements such as $\text{P}(X \leq 4) - (1 - \text{P}(X \leq 1)$.

The problems in part (b) were similar to those in part (d) of Question 1: attempts such as $\text{P}(X < 2)$ and $\text{P}(X = 2)$ were seen in some scripts. Again, many clear and correct final conclusions were provided.

Not all candidates were able to identify the initial inequality $0.9^n < 0.01$ in part (c). A variety of methods were used to solve this inequality. Some candidates used logarithms. Even here there were variations. Some candidates followed the expected approach leading to $n > \frac{ln0.01}{ln0.9}$. Almost all such candidates correctly reversed the inequality, although a few omitted this important step and arrived at an incorrect answer of 43. An alternative method involved 'trial and improvement' which worked well for some candidates. Though those using the 'Tables' often only compared 40 and 50 scoring only the first method mark in this part.

Question 4
Candidates did not approach part (a) in the manner anticipated. A correct answer of 9/20 or 0.45 was often seen, but too many final answers seemed to be $\frac{9\pi k}{20}$. This is equivalent to the correct answer but must be simplified to a numerical value in order to earn the mark.

Part (b) was well answered with most candidates correctly using the fact that the total area must be equal to 1. A few candidates used integration, starting with $\int_k^{21\text{k}} \frac{\pi}{20} dx = 1$, to obtain the correct value of *k*.

Interestingly despite having a correct expression for the variance in part (c), a significant minority of candidates did not simplify this sufficiently (often forgetting to square either the *k* or the 20).

Many correct solutions to part (d) were seen that used the standard method. Slips were sometimes seen in the initial expansion and there was some confusion between $\text{E}(X^2)$ and $[\text{E}(X)]^2$.

Other candidates used integration to find E(*A*). This was a very laborious method, and all credit is due to those candidates who persisted with all the details in order to achieve the correct final answer.

Some tried to use the formula for the area of a circle and oversimplify what was required.

Question 5
Part (a) was answered well with most scoring full marks here.

In part (b) some remained with the Poisson and some even tried the normal distribution. But the main errors were dealing with $\text{P}(X < 2)$. $1 - p$ was seen a number of times. Some candidates chose to calculate $\text{P}(X = 0) + \text{P}(X = 1)$ but often missed out the coefficient of 4. A few candidates incorrectly looked at $\text{P}(X \leq 2)$ rather than $\text{P}(X < 2)$.

There were many excellent responses to part (c). However, there was also plenty of scope for error in this eight-mark part of the question. Most candidates used the correct Normal distribution. A few ignored the instruction in the question: "using a Normal approximation" and persisted with a Poisson distribution. Most correctly remembered to use a continuity correction and many obtained the correct *z*-value.

A significant minority of candidates used a Normal approximation from a binomial distribution.

After setting up an equation, two operations were required:

- solving the equation
- 'squaring'

There were two available strategies depending on the order in which these two operations were carried out. It turned out that it is quicker and more efficient to solve the 'disguised' quadratic first and then square the numerical answer(s). Alternatively, the quadratic (in $\sqrt{x}$) can be squared first to obtain another quadratic that then must be solved. Squaring an equation required much more work. Furthermore, a substantial number of candidates performed this operation incorrectly, under the mistaken assumption that, in effect, $(a-b)^2 = a^2 - b^2$

Candidates need to make their methods clear as solutions based entirely on calculator work may not score full marks.

Question 6

Despite the algebraic nature of this question, the vast majority of candidates went on to score full marks here.

What made this question easier for candidates was the fact that answers were given allowing those who made initial errors to start again and achieve the required results.

Part (a) was virtually always correct.

In part (b) for those who managed to obtain f(x) = $a + 2bx$, this part of the question was well answered and the candidates generally achieved full marks. Occasionally F(x) was used which lost a number of marks. The other common mistake was to not explicitly show the substituting for ak and a few made a careless transcription error to lose the last accuracy mark.

In part (c) success was well correlated with success in (b) as similar principles and methods were in play. The common errors were candidates attempting to integrate xF(x), not squaring the mean and thus subtracting 6/5 and again not showing the substituting for ak.

Generally, the correct equation was obtained in part (d) when attempted and was solved correctly. Some candidates managed to form an equation in k only, which was not always a quadratic expression. Occasionally 5.2 was incorrectly selected.

Finally in part (e) the correct method was usually applied to find b and a, and where previous errors had been made there was usually correct follow through for at least one of the values. Those not making progress in parts (b) and (c) were generally still able to attempt the final parts of this question.