Examiners' Report
Principal Examiner Feedback

October 2022

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

Report on individual questions

Question 1

(a) (i) This was a nice easy start to the paper with the vast majority of students gaining B1 as they could correctly find $P(F < 3|F \sim \text{Po}(1.5))$. The few that lost this mark often did so because they found $P(F \leqslant 3|F \sim \text{Po}(1.5))$.

(a) (ii) Again, the vast majority of students gained 2 marks as they wrote or used $1 - P(F \leqslant 5) = 0.0045$. The common error seen here was to write or use $1 - P(F \leqslant 4)$.

(b) Most students were able to score marks in this part of the question. Many correctly identified Po(10) and calculated $P(R \leqslant 12) = 0.7916$. A few lost the first M mark as they found $P(R \leqslant 11)$ or $P(R = 12)$. Some students made no further progress as they failed to realise that a binomial distribution was then needed to proceed with the question. For those that had found 0.7916 and used binomial usually went on to score full marks. Those that had an incorrect probability in most cases were able to score the next 2 M marks as they used their probability correctly in a binomial calculation.

(c) This proved to be the most challenging part of the question. Some made no attempt. Others identified that Po(0.4) was needed and then either stopped or used 0.4 in a calculation for expected profit. For those that identified $e^{-0.4}$ and $1 - e^{-0.4}$ usually went on to score full marks. However, a few students lost the final 2 marks as they multiplied by an amount, often 375 which was the expected profit for 15 rolls.

Question 2

(a) Many students gained 1 of the 2 marks on offer here as they were able to sketch a graph with the correct shape. The second mark was often lost as labels were often missing, $2k$ – 0.75 was the most common missing label.

(b) This part of the question was attempted in many different ways. Those that took an integration approach were usually the least successful. In questions like this that stated 'By forming and solving an equation in k' students should be encouraged to show their working when solving their 3 term quadratic. A correct quadratic equation followed by $k = 1.25$ (which was given in the question) lost the final 2 marks.

(c) Most students realised that $\int_{0.5}^{1.25} 2x^2 - \frac{3}{4}x \, dx$ was required and went on to score full marks. A common error was to find $\int_{0.5}^{1.25} f(x) \, dx$. Those students that included $\int_{-0.5}^{0.5} \frac{1}{4}x \, dx$ usually realised that gave 0. For those that didn't, only the final mark was withheld. Students should be encouraged to show all their working including the substitution of limits in questions like these that stated 'Use calculus to find …'

(d) This proved to be the most challenging part of the question. Some students spent time finding the cdf and then did no further work. Many incorrectly thought that $Q_1 = 0.25$. Only the better students were able to find a correct equation for Q_3. A common incorrect equation seen was $k^2 - \frac{3}{4}k - \frac{5}{8} = 0$ which should have raised a concern to students as this was the same quadratic that they had found in part (b). Again, students should be encouraged to show their working when solving their 3 term quadratic. Some candidates incorrectly thought the IQR was $[Q_1, Q_3]$ and not $Q_3 - Q_1$.

Question 3

(a) Most students were able to score marks in this part of the question. Many were able to state their hypotheses in terms of p, but a few either did not write hypotheses or used no letter at all. The next M mark was gained by the majority of students as they could either write one of the probability statements required or identify a correct critical region. However, the next A mark was often withheld as both 0.0303 and 0.0173 was needed as the question stated that students needed to state the probability used which needed to be as close to 2.5% as possible. Often the final A mark was lost as only one of the two critical regions were correct or were written as probability statements. A common error was $X \leqslant 7$ instead of $X \leqslant 8$ presumably because the students had not read that the probability needed to be as close to 2.5% as possible.

(b) As this mark was followed through from the critical regions given in part (a) then it was answered well by the vast majority of students.

(c) Quite a few students used $\lambda = 7$ for their H_0 etc. As the question is testing the change in proportion the hypotheses should be written in terms of p. The majority of students used the probability method and were able to write or use $P(Y \geqslant 11) = 1 - P(Y \leqslant 10)$ to obtain 0.0985 and come to the correct conclusion. Those using the CR approach were generally less successful. A small minority of students used the binomial or the normal approximation.

Question 4

(a) It was pleasing to see that many students were able to score all 3 marks in this part of the question. The common error seen here was to use $P(X \leqslant 7) - P(X \leqslant 5)$.

(b) A range of marks was awarded for this part of the question but there were marks that were accessible to all students. The majority of students were able to identify N(56, …) so scored M1 and many of these went on to identify N(56, 33.6). A few however wrote or used N(56, 56). Many students were able to standardize = z value and so scored M1. Only a few students used z values outside the range given in the mark scheme. Those students that used a continuity correction usually went on to score full marks, however too many students failed to realise that a continuity correction was needed. Many used ± 1.98, but a few students used ±1.96. A common incorrect answer was 68 which came from using 1.98 instead of –1.98 in the standardisation.

Question 5

(a) For those students that found $d = 7$ quite often all 4 marks were awarded. For those that did not this part of the question was a non-starter. As this question was a 'Show that …' question then students needed to show their working for solving the 3 term quadratic as well as the given statement to score the final A mark.

(b) This was answered well by the vast majority of students as they were able to substitute 3.5 into the correct expression.

(c) It was pleasing to see that many students were able to deal with conditional probability and many were able to score full marks on this part of the question. A common error seen was finding $P(3.5 < X < 5.5 | X > 4.5)$ instead of $P(X > 4.5 | 3.5 < X < 5.5)$ as stated in the question. A few students correctly found $P(X > 4.5)$ but then went on to subtract this from 1.

Question 6

This question was done well by many students, only a few students did not attempt this question.

(a) Many students scored the first M mark as they were able to use $\frac{a}{a+7}$ in an equation. Many were able to set up the correct equation, however a number of students either failed to realise that there were 2 ways of getting 5 and 10 from the second bag or thought that there was 3 ways of doing this. As this question asked 'by forming and solving an equation in a' then once a correct equation had been formed then at least one further correct line of working was needed before stating the given answer. A few students did not do this and so lost the final A mark.

(b) Many fully correct answers were seen, often given in a table. Where full marks were not awarded then common errors included only having 2 of the 3 required ranges (often $R = 0$ was missed) or an incorrect probability for $R = 5$ was calculated. It is worth encouraging students to check that the sum of the probabilities = 1 or in this case that the sum of their 3 probabilities added to $\frac{193}{256}$.

Question 7

This was by far the most difficult question on the paper and part (c) proved to be challenging even to the better students. A sketch of a uniform distribution would have helped students answer this question.

(a) Many students were able to set up 2 correct equations and was often implied by $a = -11$ and $b = 29$. Some students stopped at this point as they failed to realise that $P(X > 6)$ was then required. Those that realised that $P(X > 6)$ was required usually scored full marks.

(b) Many students were able to identify that $c = 3.4$, however too many went about it in a long and convoluting manner by expanding $\frac{1}{12}(c-1)^2$ and then solving the 3 term quadratic. Many students went on to find E(Y) using their value of c. Many then went on to find E(Y^2) using their E(Y)^2, however too many students incorrectly subtracted their E(Y^2) from 0.48 or failed to square their E(Y) and so lost the final 2 marks.

(c) The final 5 marks in this question proved to be challenging and there were a large number of responses which indicated little understanding of the context of the problem. Some students left this part out completely. A few scored the first M mark as they were able to identify a correct region required. A lack of defining which distribution they were using hindered students attempts at the question. The most successful approach seen used either $X \sim$ U[10,20] with $P(X < 12)$ and $P(X > 14)$ or $Y \sim$ U[0,10] with $P(Y < 6)$ and $P(Y > 8)$.