Examiners' Report
Principal Examiner Feedback

Summer 2024

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

WST02 PE Report June 2024

General Introduction

Overall, this paper allowed all students to demonstrate their ability and knowledge of the WST02 specification. Good attempts were seen at all questions and it was pleasing to see the conclusions to hypothesis tests given in context. Questions 2(b) and 5(d) were the most discriminating part of the entire paper.

Question 1

Part (a) was routine start to the paper. Many fully correct responses were seen, but some students struggled with these straightforward inequalities.

Part (b) was a two-stage problem. Whilst the majority of students did in fact score all four marks. “Between 4 and 7 (inclusive)” was not always interpreted correctly, which resulted in an incorrect probability for the first stage. Many of these students continued on to the second stage and scored the second method mark for a correct application of the binomial distribution. On some occasions students stopped after calculating $\mathrm{P}(Q \leqslant 7) - \mathrm{P}(Q \leqslant 3)$.

Part (c) was generally well attempted. Usual errors were seen here with some attempting to calculate $\mathrm{P}(R = 10)$ or writing $\mathrm{P}(R \geqslant 10) = 1 - \mathrm{P}(R \leqslant 10)$. It was pleasing to see many students giving their answer in context.

Part (d) saw a mixed response, as is often the case when students have to give their answers in words. Many vague comments were seen that did not focus on the requirements of the Poisson distribution. Others referenced the critical region perhaps since this was an answer to a similar question from last year’s paper.

Question 2

Almost all students scored both marks for part (a) and errors seen achieving the printed answer were rare. A few students integrated the cumulative distribution function. Some students adopted a lengthier approach by attempting F(4) + (F(5) – F(4)) + (F(d) – F(5)) = 1

Part (b) was the most challenging part of this question. The numerator of the conditional probability was the issue for many students, in particular, the realisation that $\mathrm{P}(H < 1.5 \cap 1 < H < 4.5) = \mathrm{P}(1 < H < 1.5)$. Many still made a valid attempt at a conditional probability with the correct denominator.

A large proportion of students earned all three marks in part (c) with excellent attention to detail including writing the probability density function in terms of h.

A small number solved the quadratic equation in part (a), even though the question did not require this. Others used the one relevant value, $d = 10$, on the third line of the probability density function. This was acceptable, but, again, not required.

Question 3

Few candidates managed to get the mark in part (a), some referred to the stocktaking system, some just referred to all Jian's shops or stated the shops which were using the system incorrectly and some were unanswered. Many still do not have an appreciation that the sampling frame is used to select a sample and often give a partial list.

Part (b) was slightly more successful with students identifying that the shops are the sampling units.

Many were familiar with the demands of part (c) and this part was generally well attempted with most referring to the lower cost/ease as the advantage and not being as accurate for the disadvantage.

Part (d) is a routine type of question, and a majority of students were familiar with the topic. The main problem however was not completing the full demand of the question. It was not uncommon to see the correct statement $\mathrm{P}(X \leq 17) = 0.9788$, but the question requires $\mathrm{P}(X \geq 18) = 0.0212$. Some just gave the critical region without stating any probabilities. Others still give critical regions as probabilities and this is clearly not acceptable.

In part (e) many thought that Jian's belief was correct, others didn't mention Jian's belief or failed to say 20 was in the CR.

Part (f) was generally well attempted. Most worked with the correct normal approximation and of those the vast majority used a correct continuity correction. Very few missed out the hypotheses. The final A mark was lost for some not giving a fully correct contextual statement. A few approximated to a Poisson distribution whilst some simply made no attempt at an approximation and used the exact binomial distribution.

Question 4

Many failed to realise that sampling of 2 counters meant that this sample is done without replacement despite the clue given by the printed answer in part (a). Those who did the question using a method of sampling with replacement were awarded marks for a special case in part (b).

In part (b), most were able to identify the 5 possible means and there were many correct attempts at probabilities given (including for the special case). Some listed all of the different ways that P(M) = 7, for example, but never went on to amalgamate these into one probability. The mark for identifying the number of each type of counter was often missed by those who had incorrectly assumed replacement.

The overall response to part (c) was very encouraging. The mark scheme anticipated more than one method, but 'trial and improvement' was rarely seen. Successful students invariably used logarithms to solve the problem. Some had difficulty with the inequality sign and rounded down rather than up.

Question 5

A large majority of candidates scored full marks on parts (a), (b) and (c). The inequality was generally well-handled in (a), and in (b) the formula for Poisson probability was familiar to many, although some used technology for the required probability.

Part (d) was the most challenging part of the whole examination. Correct solutions were seen, but these were few and far between. Three separate (independent) events were involved:

- Jia's call is answered by the right department on the first attempt: 1 – 0.05 = 0.95
- The receptionist receives no other calls in 40 seconds: P($F = 0$) from Po(4): 0.0183
- Jia's call is answered within 40 seconds: $\frac{40-10}{50-10} = \frac{3}{4}$

Many students identified only one of the above, usually the second or third. Others, less often, stated both the second and the third. A significant proportion of students were unable to make any attempt at part (d) at all.

Question 6

In part (a) the majority of students were able to carry out the correct integration and substitute in the given limits. There were occasional sign errors in simplifying the term as $+\frac{b}{2}$ instead of $-\frac{b}{2}$.

Whilst many did put '=1' from the first step a minority did not use the '=1' until stating the final given answer and so lost the final mark. Some students over-complicated the question by first finding F(x) and using F(– 1) = 0 to evaluate the constant as '$a - \frac{b}{2}$' before using F(3)=1. The trapezium approach using areas was rare.

In part (b) there were many different approaches, often correct. The most successful method was that stated in the mark scheme, calculating $\int_{-1}^{3} x^2 f(x)dx$ and equating to $\frac{17}{5}$. Of the students who adopted this approach, most were able to integrate successfully and substitute in the correct limits leading to the correct answers of a=0.15 and b=0. Those using E(X^2) = Var(X) + {E(X)}^{2} often made algebraic errors.

The substitution of $b = \frac{1-4a}{4}$ was made at various stages; those who substituted in before integrating made manipulation more difficult. The most common error was in manipulation of algebra after substitution. Some students with algebraic errors found an incorrect value of a and then worked backwards from b = 0.1. A small number of students integrated correctly and substituted for b but failed to substitute limits, leaving an expression in terms of a and x.

Part (c) was generally well answered with many correct answers. Student who struggled to find a correct value for a then used the given value of b to find the correct value of a to use in the graph. Occasionally, an otherwise correct graph started at (–1, 0)

Those students in part (d) who used the limits $x = -1$ and $x = k$ with area of 0.2 were generally most successful. Once the correct equation had been found most students then used their calculators to solve their quadratic equation and managed to eliminate the out-of-range solution either by ignoring it or stating it was out of range. Very few students found the area of the trapezium.

There were occasional sign errors which resulted in the wrong quadratic equation being stated. Only those students who made their method clear for solving the quadratic equation were then able to secure a further mark. This question clearly stated that solutions relying entirely on calculator technology are not acceptable, so working was expected.

A common error was to treat this as a discrete random variable using the limits $(k - 1)$ and (-1).

Many students who did not score well on earlier questions were able to pick up a substantial number of marks in question 6.