Examiners' Report
Principal Examiner Feedback

Summer 2025

Pearson Edexcel International Advanced Level
In Statistics S1 (WST01) Paper 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

Summer 2025
Publications Code WST01_01_2506_ER

General Introduction

Overall, this paper allowed all students to demonstrate their ability and knowledge of the WST01 specification. Questions 6(d) and 7(e) proved to be more discriminating at the top end. Questions 2(c) and 3 showed that some students struggle with questions involving coding. Questions 1 and 4 were accessible to most students. It was pleasing to see students taking note of the instructions on the front of the paper in particular the one that says "Inexact answers should be given to three significant figures unless otherwise stated".

Question 1

This was a generally accessible start to the paper with many students scoring well on this question.

(a) This was probably the most difficult part of the question with some students still not showing clear understanding of the notation F used for the cumulative probability. Some stated it was equal to 0 whilst others believed it was equal to $\mathrm{P}(X = 3)$ or $\mathrm{P}(X = 4)$.

(b) This was generally well answered. Some took a longer route than necessary here by writing out the distribution of $3X – 3$ and $X + 2$ and comparing them individually.

(c) This was the most successfully answered part of the question with virtually all students calculating the expected value correctly. Some still go on to incorrectly divide this by 4.

(d) The instruction that working was required was heeded by most students and many went on to arrive clearly at the given answer for the variance of X.

(e) This was well attempted, however, a few students simply multiplied by –2 or failed to realise that the added constant had no effect on the variance. Some used a rounded value of the given answer and therefore did not achieve the correct accuracy in this part.

Question 2

(a) Whilst the majority of students were able to show that $b = 43.4$, a few failed to show sufficient working to gain full marks. Students need to be more thorough in presenting all steps in their working to ensure that they gain full marks in 'show' questions. Some students used $\frac{24\times 47+30b}{54} = 45$ and just wrote $b = 43.4$ without providing at least one line of intermediate working.

(b) Some students calculated the variance but not the standard deviation. Others were able to establish the fraction $\frac{66876+73826}{54}$ but then subtracted 45 instead of 45^2.

(c) It was rare to see fully correct solutions to his question. Many students failed to give supporting reasons and therefore scored no marks here. In part (c)(i) many mentioned that the variance would not change because the data was coded, which is of course insufficient. The key or the critical point here is that variance is unchanged because the "coding" only involves addition. Part (c)(ii) was more often correct, as all that was required was a statement that the total score had increased, some incorrectly went on to say that the overall mean increased by 2 marks. Part (c)(iii) was the most challenging: many simply stated that the variance would not change, presumably only referring to Class B rather than the overall variance. It was rare to see students reason that the means would be closer together so there would be less overall spread in scores.

Question 3

There were very few blank scripts, but a considerable number of students lost marks in parts (a) (c), (d), (e) and h(ii).

(a) Among the more common unacceptable comments were "statistical models give more accurate/precise results" or saying it was quick, cheap, easy or to make predictions. The most common correct answer was to simplify the real world.

(b) This was generally extremely well done with only a handful not gaining this mark usually for $10036.45 - \frac{339.25}{12}$ or for $10036.45 - \left(\frac{339.25}{12}\right)^2$.

(c) This was a challenging part of the question for many. Coding for this was not considered by many students who spent much time and effort working out various quantities such as $\sum td$ or $\sum d^2$ which sometimes, but rarely, led to the accurate answers. Very few knew to use $S_{td} = \frac{-91.55}{11.5}$ or $S_{dd} = \frac{445.57}{11.5^2}$. A common error was to use $S_{dd} = \frac{445.57}{11.5}$.

(d) Most used their answers from (c) to gain at least 1 mark. Few realised they could use $\frac{S_{tw}}{\sqrt{S_{tt}S_{ww}}}$, although those that did were more likely to score the accuracy mark, and many failed to appreciate that $-1 \leq r \leq 1$. Even those who scored full marks in (c) often lost the accuracy mark by giving their answer as 0.84 or reaching –0.843 from correct but rounded values for S_{dd} and S_{td}.

(e) As usual, students find written expression in context rather difficult. Too many students just described the correlation rather than giving an interpretation in context. A number gave a correct interpretation even though their r would suggest a positive relationship. An invalid value of r in (d) rightly precluded this mark for many. Some interpreted their value as a gradient rather than a correlation coefficient.

(f) Of those who did attempt this part, most scored full marks. Few did not show the required substitutions, and the means were almost always substituted the correct way round first time, although some did lose the final mark for not stating the given equation.

(g) In this part a common misconception seemed to be that the equation stayed the same, others tried to recalculate the equation using S_{td}, S_{dd} and S_{tt}, but many scored the mark by simply replacing w with $11.5d$.

(h)(i) Generally, those who had (g) correct also had this part correct. Although some used insufficiently rounded values or gave an answer of 2.1.

(h)(ii) Many incomplete responses were seen to this part. A general interpretation was often seen i.e. as temperature goes up gas consumption goes down, or most commonly missing off the units. Decrease by –0.3 (m^3) was seen fairly frequently.

Question 4

(a) Many correct answers were since here. Of those making mistakes, many incorrectly assumed independence.

(b) Again, many correct responses were seen here. Even those making slips in part (a) were able to gain a method mark for an answer 0.14 greater than their part (a). Common incorrect answers assumed that C and D were mutually exclusive and hence thought that $P(C) + P(D) = 0.59$.

Question 5

(a) Most student were able to correctly identify the interval $\mathbf{25 \leq x < 30}$ as having the highest frequency. However, a significant number failed to identify the second-highest interval, $\mathbf{20 \leq x <}$

25. Though many students made an attempt at using frequency density, many did so for the wrong bars.

(b) While most students were able to reach the correct value of 45.8, many failed to round it appropriately to an integer (45 or 46). A notable number of students either didn't round at all or rounded incorrectly. Some students misused interpolation or confused values (e.g., getting 1.2 instead of 0.8).

(c) Despite the familiarity of this type of question, there remain a large proportion of students who struggle to apply linear interpolation accurately. A common issue was incorrect or incomplete setup of the required expression or equation. While some students managed correct working, a number of them substituted incorrect values such as 114 instead of 144.

(d)(i) Errors here mirrored those seen in part (c). A frequent mistake was to calculate 3×25.14 rather than apply the correct formula or method.

(d)(ii) This subpart was generally answered better. Most students understood the need to compute Q3 – 25.1, using the given lower quartile. However, some recalculated Q1 unnecessarily or substituted a formula for detecting outliers instead of using the interquartile range.

(e) Students generally knew how to approach this comparison task. While many were able to describe the skew correctly, a significant number made unsupported or incorrect conclusions, such as identifying the distribution as negatively skewed despite evidence to the contrary. Some referenced the mean without attempting to calculate it, and others described the shape without justification.

(f) This was generally well answered. Most students correctly referred to the symmetry or skewness of the distribution. A few incorrectly described the data as "uniformly distributed" Some simply ignored what part (e) was showing and stated that the data was continuous which is not a sufficient condition here.

Question 6

(a)(i) Almost every student recognised the need to demonstrate that they were subtracting the mean of 502 from the 508 and dividing the result by the standard deviation of 3. A very rare error was that of dividing by the variance. Very few students relied solely on their calculator and those that did scored no marks. On some occasions students went on to incorrectly calculate 1 – 0.0228.

(a)(ii) Most students successfully found a method using their value in (a)(i) to obtain these marks. Drawing a diagram helped many. Most went down the route of calculating P(X < 508) – P(X < 496). A few students recognised the symmetry and calculated 2 × (their answer in (a)(i) – 0.5).

(b) Students struggled to put a full solution together here. Many failed to recognise the need to find an expression in terms $p^2 \times (1-p)^2$ (with multiplication by 2 often seen rather that the powers of 2). Some manually worked out that there would be 6 combinations and some used $^4C_2$ but many used 4 as a multiplier or no multiplier at all and fewer than half the students were able to arrive at the correct answer.

(c) A large number of students were able to recognise the need to equate $\frac{1038.51-1024}{\sigma}$ to a z-value, and the vast majority were able to secure at least one mark. Many students truncated or rounded to give 1.03, 1.04 or 1.036 failing to give the statistical tables accuracy of 1.0364.

(d)(i) Many students engaged with this trickier part and most recognised the need to standardise k with 502 and 3 and $2k$ with 1024 and 14. Most then attempted to set up an equation though many did not recognise the need to make one of their expressions negative in order to maintain compatible signs. A rare error was made by those recognising that they should not simply equate their standardised expressions but using 1 – for one of them rather than a multiplier of –1. Most students were able to successfully rearrange their equations to arrive at a value for k however the incorrect 495.5 was often seen.

(d)(ii) Those with a value found in part (d)(i) were almost always able to standardise appropriately with their value of k securing the method mark. Those with the correct value from part (d)(i) almost invariably went on to score both marks here.

Question 7

(a) This part was accessible to the majority of students, as it involved a basic probability calculation using a tree diagram. Students who were familiar with the concept of probability without replacement generally executed the task correctly. One common error was a misapplication of replacement probabilities, where students did not adjust the total number of beads after each selection. Some students appeared to construct their tree diagram with frequencies rather than probabilities, and usually accrued few marks for subsequent work.

(b) This part was generally well done by the vast majority of students who had correctly completed part (a), as it relied on interpreting the tree diagram to identify the scenarios that result in exactly one yellow bead. Most students understood the concept of "exactly one" and were able to answer this part successfully. Some added an extra product, presumably confusing “exactly one yellow” with “at least one yellow”. Some only included one of the two possible combinations.

(c) This again was a fairly straightforward for most students. Most use a methodical method by listing all combinations and adding together the required products. Only a small minority of students opted for the more efficient 1 – method by calculating the probability of selecting two yellow beads and subtracting from 1. In general if the tree diagram was correct most students scored full marks here.

(d)This part presented a moderate challenge, as it required students to apply the concept of conditional probability with a significant proportion of students still unable to distinguish between this concept and that of independence.

(e) This was the most demanding part of the question, and indeed the paper for many, and provided strong differentiation for the highest achieving students. This part required students to translate the probability model involving beads into an algebraic expression, and then manipulate this to form a quadratic equation. Students who were comfortable with combining algebra and probability often managed to progress to the correct form of the equation, but many struggled with the algebraic manipulation required to arrive at the given form. Those who were able to successfully cancel factors before attempting to simplify were generally far more successful than those who were not able to do so. Many were able to earn at least one mark for correct denominators and equating the products to 3/235. Correct numerators were seen less often. Some just left their answer in an unsimplified or non-polynomial form.

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