Pearson

Examiners' Report
October 2025

International Advanced Level in Statistics S1

WST01/01

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 learners at: www.pearson.com/uk

October 2025
Question Paper Log Number P84842A
Publication Code WST01_01_2510_ER
All the material in this publication is copyright
© Pearson Education Ltd

General Introduction

Overall, this paper allowed all learners to demonstrate their ability and knowledge of the WST01 specification. Questions 2(d), 5(c)(ii) and 8(c) proved to be more discriminating at the top end. Question 1 was accessible to most learners. It was pleasing to see learners 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 learners scoring well on this question.

(a) Most learners were able to obtain all three marks. Where errors were made it was very rare to find learners unable to find the correct method for at least one of the parts. The most typical error was those learners who incorrectly applied the formula for S_{hw}.

(b) Again, the vast majority of learners were able to substitute their values into the correct equation to gain the method mark. A few accuracy errors were made with 0.95 seen on some occasions.

(c) Almost all learners were sufficiently aware of the method required to score both method marks. A small number lost the first mark for not using the given value of 286 in their denominator, sometimes attempting to calculate S_{hh} themselves. A very small number of learners used a different quantity in their denominator but in general it was rare for learners with correct values in part a) to fail to work out where 3.27 came from. Most learners took note that it was a 'show that' to demonstrate their method to find **b** and **a**. The requirement for maintaining full accuracy in their calculations is still a point for learners and centres to work on however. It was more common than not for learners to lose

their final accuracy mark by obtaining 45.85 for their value of **a** having used the rounded 3.27 for **b**. Learners either failed to notice, or chose to disregard, that this value did not round to 46.0. Some evidently did notice and chose to round their answer to 46 but this was not sufficient to score the final mark. Some failed to use the means of **w** and **h** and on occasion these were misplaced. A more common though rarely seen error was made by learners failing to write the required equation having calculated their **a** and **b** or those who gave their final answer in terms of **y** and **x**.

(d) While there are still a sizeable number of learners who are unable to interpret their gradient attempts were better than in previous sessions. However, some chose just to describe positive correlation without context or specific values, while others made errors with either units, or by switching the variables in their explanation and incorrectly stating an increase of 3.27 in height.

(e) Almost all the learners were able to score both marks for substitution into the given regression line. Almost all learners used the given equation even if it did not match their own. Where errors were made it was by those who did not work with sufficient accuracy.

(f) While most learners were aware of the requirement of the values being within range as correct reasoning many were unable to describe clearly that it was the heights that needed to be in range since the regression line was **w** on **h**. Most learners identified that the first estimate was the more reliable one but often gave reasons which lacked the necessary clarity either by stating that the weight was in range or just saying that **it** was in range. A surprisingly small number of learners used interpolation as their justification for selecting the first estimate, though a few more stated extrapolation as their reason to reject the second estimate. Of those learners seemingly unaware of what made for reliable estimates most wrongly chose the second estimate since it rounded more easily to a whole number or had fewer decimal places.

Question 2

(a) This part was generally very well done with wrong values for the quartiles rare.

(b) Most learners showed their working, even with incorrect quartiles, to gain the first two marks, but some either did not state there was one outlier of 139 or made an incorrect comment such as "69.5 is an outlier because 69.5 < 72". A few failed to realise they actually had more than one outlier from incorrect quartile values.

(c) This was less confidently answered. Almost every response earned the first M1 but far too many drew the lower whisker to the outlier limit which is beyond the given data and so lost the second M1 and hence the A1. Most did score the third M1, with few miscalculating the median value or incorrectly plotting their quartiles. Occasionally responses that scored the 3 M marks, lost the A mark by omitting the outlier.

(d) This proved to be the most discriminating on the paper. The majority of responses did state positive skew (rather than negative) but it was very rare indeed that this was correctly interpreted in context. Many probably thought that use of $Q_3 - Q_2 > Q_2 - Q_1$ was the description and "positive" was the interpretation. Many incorrect comments about there being more above the median than below were seen.

Question 3

(a) Most learners seemed aware that a probability of 1/6 was involved but some may have got confused by the request for a probability function and over complicated it. The most usual correct answer was learners writing a probability distribution in a table. It was evident that learners were not sufficiently aware of correct notation. A small number gave the cumulative distribution instead of the probability distribution. A less frequently seen complete method was demonstrated by those describing $f(y) = 1/6$ with the correct values then stated.

Some learners remembered to say that $f(y) = 0$ for all other values. Many learners however simply said 1/6 without referencing the Y values at all.

(b) This was perhaps slightly better done than in previous series and many learners included uniform in their description. Very few precise wordings of discrete uniform distribution were seen but the mark could be scored for reference to uniform alone as long as not in a continuous context. There was considerable uncertainty about what was needed here though and while the most typical wrong answer was discrete random variable frequently seen alternatives were even, equal, discrete, continuous or normal.

(c) Most learners were aware that they needed to multiply by 4 and add 3. Many wrong answers were seen involving $4 \times P(Y = y) + 3$ instead of $4 \times E(Y) + 3$. A minority of learners chose to find the new variable $4Y + 3$ but were not always successful in multiplying their new variable by the probability. Those attempting to find $E(Y)$, including some who hadn't realised that a probability distribution was an acceptable answer in part (a), usually found $E(Y) = 3.5$. Both use of symmetry and calculation were seen. Those with a value of 3.5 were usually able to secure all three marks. Some though believed they needed the variance or simply multiplied $E(Y)$ by 4^2.

Question 4

(a) The opening part of this question was generally well answered with the two simultaneous equations stated and solved accurately. For those who failed to score 3 marks most derived the equation involving E(*X*) but surprisingly failed to give the equation involving the sum of probabilities = 1. Where there were errors in solving, it often involved the evaluation of $0 \times 0.1 = 0.1$.

(b) This was generally well answered with a table the most common presentation. Those who opted for a simple list of cumulative probabilities often lost the marks as they failed to associate each with the appropriate *x* value.

(c) Those who realised that this was equivalent to $\mathrm{P}(X < 0.5)$ were able to give the correct answer, even if they had not completed the cumulative distribution in (b). A common wrong answer was 0.85.

(d) This part was also well answered with the most successful learners clearly applying the correct formulae $\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$ and $\mathrm{Var}(5 - 3X) = 9 \times \mathrm{Var}(X)$ with errors typically due to arithmetic inaccuracies rather than misunderstanding of the underlying concepts. Of those who did not gain full marks, most were able to gain the first M1 by showing the method to find $\mathrm{E}(\mathrm{X}^2)$ with three correct non-zero products, even if the lack of brackets when squaring −1 meant that they arrived at the wrong value of 4.5 rather than 4.6. The majority of responses did demonstrate clear use of $\mathrm{Var}(X) = \mathrm{E}(X^2) - [\mathrm{E}(X)]^2$ using their value of $\mathrm{E}(\mathrm{X}^2)$, but the correct application of the transformation $\mathrm{Var}(5 - 3X) = 9 \times \mathrm{Var}(X)$ was less common with some just multiplying $\mathrm{Var}(\mathrm{X})$ by 3 and others multiplying by −3 and adding 5. A few did attempt to calculate $\mathrm{Y} = 5 - 3\mathrm{X}$ and $\mathrm{Var}(Y) = \mathrm{E}(Y^2) - [\mathrm{E}(Y)]^2$ find but these were rarely successful, again often due to arithmetic errors dealing with negative numbers.

Question 5

(a) Many were able to successfully score all three marks here which was pleasing. However, a number still seemed to have little understanding of what was required. Less confident learners were often able to show the relevance of the frequency densities to secure the M mark (often seen by the table) but were often not able to understand how this helped with the answer. Those obtaining the correct width of 7.2 were often successful in gaining all three marks but a typical wrong answer was to have a width of 3.6 instead. Many of these learners did recognise however

that the area of their new bar should be 11.52 to enable the method mark to be earned.

(b) Learners are familiar with the requirements of this question and the majority demonstrated familiarity with the method. A smaller number worked back from 140 but usually successfully. Those learners without the method rarely showed much appreciation of linear interpolation and answers were usually either correct following the correct method or displayed no idea of how to approach it. Some realised they needed to equate two ratios but made numerous errors in attempting it.

(c)(i) The vast majority of learners were able to find the mean of y and code it correctly to gain a value of 240 for the mean of w. Some however used a mean of y as 6000 rather than 6000/300 or attempted to use a formula to calculate the sum of w without success.

(ii) Less certainty about coding was displayed with regards to the variance and answers of just doubling their variance or doubling and adding 200 were frequently seen. Dividing by 4 rather than multiplying was also a common error. Some also believed that Var(w) was equivalent to Var(y). An alternative successful method came from those calculating the standard deviation to be 10 doubling it and then squaring their answer. A very small number attempted to find sum w^2 but without success.

Question 6

(a) Although the majority of responses scored full marks, some learners fail to recognise that independence between **A** and **B** means that $\mathrm{P(A \cap B)} = \mathrm{P(A)} \times \mathrm{P(B)}$. Of those who did, poor algebraic notation and manipulation, particularly lack of or misuse of brackets, often caused problems. $\mathrm{P}(B) = 0.33x$ rather than $0.33 + x$ was common. Others failed to correctly identify the regions which made

up $\mathrm{P}(A)$ or $\mathrm{P}(B)$. It was concerning to still see values for *x* outside [0, 1] being used in further work without any worry.

(b) Performance in this part was more mixed, those who failed to make headway in (a) rarely scored anything here. Learners were expected to use the given conditional probability $\mathrm{P}(B \mid C') = \frac{31}{75}$ applying the definition $\mathrm{P}(B \mid C') = \frac{P(B \cap C')}{P(C')}$. Those who carefully identified which regions corresponded $B \cap C'$ and C' to and were able to form a correct equation and solve for *z*. A number of learners omitted to subtract the regions C of correctly when calculating $P(C')$, or used the total probability of rather B than $B \cap C'$, leading to incorrect results.

(c) This was generally well answered with most responses using the sum of probabilities = 1. Those who had correct or credible values for *x* and *z* usually scored both marks due to the follow through, although some lost the A mark if their values led to an impossible value for *y*. This did not seem to worry them, and it was rare to see attempts at correcting earlier work.

(d) This tested learners' fluency with set notation.

For part (i), successful learners correctly interpreted $(B \cup C)' \cap A$ as "the part of A that lies outside both B and C" and selected the appropriate probability region from the Venn diagram. For part (ii), those who understood as "the parts of A or B that are outside C" were able to sum the corresponding probabilities accurately. Common errors included adding or subtracting incorrect regions due to misinterpretation of the set symbols, particularly confusing union and intersection. Generally, learners who were successful in parts (a) and (b) usually gained full marks in (d).

Question 7

Overall, the topic of the normal distribution is still poorly understood and unlike the other questions entirely blank answers were frequently seen.

(a) Learners mostly had little difficulty in demonstrating their knowledge of standardising and the first mark was usually gained successfully. Seeing division by the variance was seen but only rarely. There was little evidence of overreliance on calculators to obtain results, though a number were unclear about how to find the correct probabilities from their standardised values. A typical wrong answer seen was to find 0.9772 – 0.9332 rather than 0.9722 – (1–0.9332).

(b) Learners were not clear about the use of conditional probability, and many were unable to understand the ratio of probabilities required. Many failed to appreciate that a ratio was required at all and there were several scripts believing that finding $P(W > 53.55)$ was all that was required. While learners were able to gain credit for understanding the need to divide by their answer in (a), most did not realise the numerator that was required.

(c) Many learners were able to gain marks here despite having struggled earlier in the question. The typical misconception remained those who equated their standardised values to probabilities, either $0.3, 0.9, 0.1$ or $P(Z < 0.3)$ and $P(Z < 0.1)$, but that was perhaps seen less frequently than in previous series. Those with z-values mostly recognized the need for 4 decimal place accuracy to match the values used in the statistical tables provided. Incompatible signs remain a common problem for learners.

Question 8

(a) This part was generally well attempted with most responses recognising the required probability of being late is obtained by summing the products of each route probability with its corresponding conditional probability of lateness. Clear

layout, writing each product explicitly before adding, was a consistent feature of correct responses. Occasionally 1 – 13/15 or other "fudged" working was seen in weaker responses.

(b) This was more discriminating but those who set their working out clearly in (a) were generally able to show the required working for the conditional probability in (b). Again, there were a small but significant number of attempts which incorrectly arrived at the given answer. There appeared to be a difficulty in relating the conditional probability in a tree diagram with conditional probability calculated algebraically.

(c) The unusual nature of the tree diagram in part (c) caused problems for many, especially for those who could not calculate/relate to the conditional probabilities, even if they had successfully answered part (b). Most however, were able to gain the first B mark for 13/15 and sometimes the second mark. Those responses which showed clear working, including checking probabilities added to one on each set of branches, before attempting to complete the tree diagram generally fared better.