Pearson Edexcel

Examiners' Report
Principal Examiner Feedback

January 2025

Pearson Edexcel International Advanced Level
In Statistics S1 (WST01) Paper 01

January 2025
Publications Code WST01_01_2501_ER

General Introduction

Overall, this paper allowed all students to demonstrate their ability and knowledge of the WST01 specification. Students should be advised to read questions carefully as marks were dropped carelessly for not meeting the required demands. In particular questions that ask a student to show something is true require all the steps in the working to be shown. Questions that state 'Solutions relying entirely on calculator technology are not acceptable' require students to show their method and not just state values from their calculator. These are usually used when standardisation is required or when solving simultaneous equations. Students would be advised to take 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

Part (a) saw mixed success for students as some were unable to name the probability distribution correctly. Some students simply described the event or gave the probability distribution rather than naming it. Simply writing discrete, or random, or writing discrete random were common errors and some invented names such as equal or fair or non-continuous. Some even chose normal as the only distribution they could name.

Most students were able to gain the mark in part (b) though 0.25 was the most typical incorrect answer seen.

Again in part (c), students confidently displayed their knowledge here. Many did not appreciate that symmetry could be used to find the mean, instead using the formula. Only a very few divided their answer by 4 in error, or used just the values without their probabilities.

Working was required in part (d) so it is important that students show all stages in order to gain full marks. The vast majority of students were able to achieve all three marks. Of those failing to gain the marks the most typical errors did involve believing Var(B) was E(B^2) or forgetting to square their E(B) value when subtracting it from E(B^2), despite having written the correct formula. Others attempted to divide either or both of their E(B^2) or [E(B)]^2 values by 4.

In part (e) given that the vast majority of students recognised the need to find combinations of R and B that would sum to 4 it was surprising how many of them were unable to do this successfully. While most were able to identify at least 4 correct combinations the most typical errors were to presume that the reverse of each combination such as (1, 2) and (2, 1) were equally applicable and failing to take on board that only the 1 and 3 values appeared on both dice. Some didn't write down any combinations at all and simply gave the correct probability calculation $6 \times \frac{1}{4} \times \frac{1}{4}$ leading to the correct answer to gain full marks. Though the advice is that students should always show sufficient working to make their method clear. Students using a sample space diagram were almost always successful.

In part (f) the majority of students were able to identify at least 3 correct combinations and it was not untypical for students who were unsuccessful in part (e) but to be successful in part (f). A small number simply gave the correct probability calculation $5 \times \frac{1}{4} \times \frac{1}{4}$ leading to the correct answer to secure all the marks.

Part (g) was a good discriminator at the end of this question. Students got confused by the cumulative distribution function and typically tried to apply their knowledge of this too soon in the process. The most typical error was to attempt to find P(D = 4) or P(D = 5) in terms of p and then summing the probabilities to 1 but since this should have resulted in the p's cancelling out this got them nowhere. Those recognising that they could find the difference of 4 (5, 1 and 7, 3) or 5 (7, 2) from the combinations on the two dice were usually successful in finding P(D = 4) or P(D = 5). They were almost invariably successful in then correctly labelling their probabilities or giving a probability distribution. A few students believed that they could find P(D = 5) from $1 - \frac{1}{16}$ not appreciating that this was in fact P(D = 6).

Question 2

In part (a) the majority of students knew that the range is the smallest value to the largest value, but strangely did not subtract the values to find the difference, instead leaving their answer as eg $0.21 \leq x \leq 0.87$ or 21 – 87, losing the mark.

Part (b) was relatively well answered with the most common error misreading the key and giving an answer of 48.

Issues with the key caused the most errors in part (c), with many giving their answer as 66 or 0.66. There were also cases of the equation being solved as though "a" was a multiple rather than representing the decimal in the data value. eg Correctly getting to $0.6a - 0.35 = 0.31$ but then "solving" to find a=1.1

Part (d) was almost universally correct. However, some quoted the formula but did not show the values they had substituted so lost marks.

In (e)(i) while most were able to gain the first mark (usually for showing 65×0.204 or 32.63), the second mark was often lost for not showing intermediate steps of working to find the given answer.

In (e)(ii) fully correct solutions were rarely seen showing a lack of understanding of combining summary statistics by many students. Common errors were 13.4228 missing from the numerator, using a denominator of 18 or 47, or using 8.91/65 instead of 0.502.

<u>Question 3</u>

Part (a) was generally very well attempted by all students. A rare minority of students multiplied the two probabilities together to obtain the probabilities for the second set of branches.

Most were able to earn the method mark for using the correct products from their tree diagram in part (b)(i) and indeed many fully correct solutions were seen.

In part (b)(ii) the conditional probability required is more discriminating. Many only calculate the probability required for the numerator. It was disappointing to see that a significant number of students who had a correct express gave their answer to only 2 significant figures.

Question 4

Part (a) of this question was typically completed successfully with the majority of students accurately completing the Venn diagram. Common mistakes included not having the 14 on the outside of the diagram or the numbers in their diagram totalling 300 rather than 100. Without the '14', students could still proceed with parts (b) and (c) without any problem. A fair number of students put probabilities on their Venn diagram.

Whilst part (b) was reasonably well attempted, some students cannot set up what is a standard proof properly failing to write out and label P(A), P(C) and P($A \cap C$) losing both marks or did not state which test they were using thus not gaining the second mark. Others who knew which test they needed to use to show independence failed to score the method mark by using the numbers from their Venn diagram rather than converting them to probabilities.

The vast majority of students were able to score the mark in part (c)(i) even if their Venn diagram was incorrect. Most were able to correctly answer (c)(ii) realising the space is reduced to 50 from which to choose. Of those who struggled with the conditional probability the common mistakes included having an incorrect denominator or just having an answer of 12/100 or 12/50.

Question 5

In general students were able to successfully standardise and get 0.885 for part (a)(i), however, several students then went on to do 1 – 0.885 and lost the final mark. A few failed to show the standardisation explicitly asked for and lost both marks.

In part (a)(ii) there was a disappointing lack of understanding of the phrase 'correct to the nearest 50kg' and students rarely used $675 < S < 725$. The vast majority opted to use $650 < S < 750$ or left this part blank completely. The mark scheme allowed students who did not understand the bounds to still score 4 out of 5 marks through the SC if they demonstrated their understanding of the normal distribution. However, again some failed to score all 4 by not showing their working fully to gain the method marks. Students should be reminded not to rely on calculator technology when the question explicitly states that standardisation needs to be shown.

In part (b)(i) students were often able to correctly find 1 correct equation, 680 – mu/ sigma being the one most commonly used here. The other equation often had incompatible signs or insufficient accuracy for the final A1; use of the percentage points table is still problematic. Students using calculators should show this value to an equivalent degree of accuracy as to those in the tables (ie 4 decimal places). If students did have the two correct equations in part (b)(i) it was common for them to get the two correct values in (b)(ii). However, it was clear that some students had used calculator technology to solve the incorrect equations they had found, hence missing out on the method mark in part (ii) for showing their working in attempting to eliminate a variable.

Question 6

In part (a) most gained full marks by correctly calculating S_{tt} and then substituting into the formula for *r* (some did this by calculating S_{tt} within the calculation of *r*). Occasionally marks were lost by not giving their answer to 3 significant figures.

Part (b) was more challenging, and it was rare for students to gain both marks here. The most common issue was mentioning the graph in terms of correlation rather than how the points would be seen. Many identified that "as *w* increases, *t* decreases". Fewer mentioned that the gradient would be negative. The least common correct answer seen would be to state that the points would lie close to the line of best fit. Some incorrectly thought that the negative correlation indicated that the graph should be non-linear.

Part (c) was very well attempted with correct and accurate methods used. The most common error was to find the values of *b* and *a*, but to omit the equation for the final the mark. Some did not give these values to the required 3 significant figure accuracy. There were a number of students with sign errors when calculating *a* and some who mixed up the variables.

Part (d) saw a mixed response but on the whole there were some promising attempts. Some misinterpreted the result as a double negative stating that the time "decreases by –0.688". Some again mixed up the variables in their numerical interpretation of the gradient.

Most students usually gained at least one mark in part (e) here but it was rare to gain all three. Part (i) was the most common correct part. It did seem that many students had guessed answers here.

Question 7

In part (a) many students struggled to imagine the usual grouped frequency table with which to apply their knowledge of linear interpolation and did not recognise that the quartiles on the box plot represented frequencies of 50, 100 and 150. Those that could combine the scale on the box plot with the frequencies represented were usually successful in forming of appropriate ratios to secure both marks. Some students (usually successfully) used percentiles rather than frequencies. A few students correctly obtained the value of 110 cabbages estimated as weighing less than 570g but then forgot to find the probability.

In part (b) almost every student was able to substitute their values to access the mark here. A few however were uncertain that this was all they had to do and tried to rethink their answer having calculated the correct value of 950 leading to a number of alternatives such as 951 or more typically 1000. Some students found the language of 'least' and 'minimum' in the question difficult to decipher sometimes resulting in students subtracting from Q1 rather than adding to Q3.

Most students were able to recognise in part (c) that since the box plot was symmetric then a normal distribution was supported. Most justified this by acknowledging that Q3 – Q2 = Q2 – Q1 rather than stating that the distribution was symmetric but this was sufficient to secure the marks. Some gained the mark for stating that there was no skew, but many lost the mark for failing to provide a reason for their support of the normal distribution or for providing an appropriate reason but forgetting to actually say that this supported its use. A few suggested that it was not an appropriate distribution to use for a variety of reasons with some distracted by the outlier and believing that this made the use of a normal distribution inappropriate. A few wanted to justify the use of the normal distribution by saying that the mean was equal to the median but this did not score the mark since the mean value was unknown.

In (d) the most typical responses were either to leave it blank altogether or to gain both marks. Those processing the information to write or to recognize it was $P(W > 2\sigma)$ quickly saw that this led to simply finding $P(Z > 2)$. Some however believed that they needed to find the value of the standard deviation and got nowhere as a result. Some students spoiled an otherwise good approach by subtracting their value of 0.0228 from 1 and a few students failed to accurately read the question and presumed that they would need to multiply their answer by 2, unfortunately costing them the final mark. Another typical error seen was to overlook that $Z = 2$ and use 1.96 instead.

The final part of this paper was accessible to many students and despite earlier errors in this question, many were able to make progress here recognising that they simply needed $560 + 2\sigma = 1000$ to obtain the correct value of 220. Some failed to realise the continuous nature of the data opting for an incorrect answer of 219.