Pearson Edexcel

Examiners’ Report
Principal Examiner Feedback

January 2024

Pearson Edexcel International Advanced Level in
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

January 2024
Publications Code WST01_01_2401_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. To make methods clear, stating formula before substitution is always beneficial. 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".

Report on Individual Questions

Question 1

Virtually all students made a strong start to the paper with very few errors seen in part (a). Of those who made minor slips in calculating the first frequency, they were able to score the method mark for frequencies that summed to 104. Some opted to label the frequency density axis, and this was often followed by two correct frequencies.

Part (b) also saw a high success rate with most obtaining a correct fraction and adding it to 13. On some occasions students used 204.5 and this is allowed as an acceptable method. Mistakes, when seen, were generally found in the denominator of the fraction.

Despite the values being very close together, most commented that since mean < median, there was negative skew. Of those who went on a quartiles approach, they were often calculated correctly and shown so the mark could be awarded – though students should realise that this involved a lot of unnecessary work for only one mark.

Many scored one out of the three marks available in part (d) for using interpolation to find 220 (and gave this as a fraction over 408) not reading the question carefully enough to realise the demand asked for the probability for **both** plants. Incorrect attempts at finding 220 usually came from not dividing 32 by 4 or 120 by 2 so ending up with more than 220. This in turn meant their number was outside the range allowed for the second mark. Of those scoring full marks, a repeated fraction of $\frac{220}{408}$ seemed more common than the lack of replacement approach.

Question 2

In (a)(i) the overwhelming majority calculated the mean correctly. Most gave as a rounded decimal, while some left it as a fraction, both of which were acceptable.

The majority substituted values correctly into the formula to find the given answer 8.57 in part (ii). However, many lost the accuracy mark through not showing an intermediary step or a more accurate value before the given answer of 8.57. It was common to see 71.8 or 71.83 (rather than 71.833 or better) used as the mean in the calculation for standard deviation, which does not lead to an accurate enough answer in a "show that" situation.

The calculation of the required summary statistic followed by the calculation of the product moment correlation coefficient was accessible to nearly all students with the vast majority scoring full marks in part (b). Occasionally an answer of 0.85 was seen, so lost the final mark for not giving the answer to the required accuracy.

The first part of (c) was often well answered as the effect of coding on the mean is generally well-understood. However, for the standard deviation, many students forgot that subtracting a constant from all data values has no impact. It was common to see students subtracting 32 from 8.57, leading to a negative standard deviation which should have caused students to realise a mistake had been made. Another common mistake was to square the $\frac{5}{9}$.

Most were able to score at least one mark in part (d) as students recognised that the PMCC did not change, and either stated this in words or gave the same numerical answer as in (b) but many were not able to give an acceptable reason why this was the case. A typical misreading of the question meant that some tried to interpret the value of their *r* value and comment on the strength of the correlation.

Question 3

The completing of the tree diagram in part (a) was done well.

In part (b), most were able to set up an suitable equation in terms of *p* and solve this correctly.

Even if errors in part (b) occurred, nearly all were able to obtain follow through marks in part (c) for multiplying their value of *p* by $\frac{1}{8}$.

Part (d) caused the most difficulty for students. A common incorrect solution seen was $\frac{0.4}{0.89} = \frac{40}{89}$ failing to understand that the numerator should be the product of the two probabilities.

Question 4

Most students were able to find at least one of the quartiles required in part (a), but quite often they did not order the Spanish data and simply selected the third item from the list leading to a common incorrect value of $Q_1 = 24$. Most went on to earn the method mark by using the outlier expression with their values. With those finding correct quartiles, nearly all went on to give a suitable conclusion showing 90 as an outlier.

Students should be working to a sufficient degree of accuracy such that the final answer can be found correct to 3 significant figures. There were a number of cases in part (b) where students rounded the value of the gradient of the regression line and therefore did not find an accurate value of the intercept. Incorrectly using $n = 11$, instead of $n = 10$, was another common error. Occasionally students forgot to write the given answer at the end and sometimes this was stated in terms of y and x instead of s and f.

Some pleasing responses were seen here in part (c). However, some still do not understand that interpretation should always indicate context is required. Sometimes 'mark(s)' was omitted and sometimes an oversimplified answer of 'positive 'gradient' was given. Some inverted the interpretation and thought 1.04 was the increase in French marks. Others incorrectly involved the –5.72 as part of the interpretation.

Part (d) was the most accessible part of the question with the majority gaining full marks. A few students mistakenly took 55 and 18 to be the Spanish marks and not the French.

In part (e) most knew that the estimate from (d)(i) where f=55 was the more reliable. However not all could adequately explain why. They needed to refer to the range of values of the French marks when commenting about the range. Some better responses made it clear which estimate involved interpolation and which involved extrapolation.

Question 5

Part (a) was answered well by the majority of students and it was pleasing to see that students realised that, as this was a 'show that' question, the standardisation was needed to be shown.

Part (b) was one of the most discriminating parts on the entire paper since many were unable to correctly work out the probability that an athlete qualifies for the final. Many just used 1 – 0.3821 as this probability in the denominator of the conditional probability. Most were able to access at least two marks in this part for finding the probability that a throw was more than 44 metres. Some were able to score an additional mark by setting up a ratio of probabilities with a correct product of probabilities on the numerator and what they believed to be the probability of qualification on the denominator – provided this was not 0.6179. It was pleasing to see those students working with sufficient accuracy throughout and arriving at a correct answer to the required degree of accuracy.

Question 6

Most made a good attempt and managed to show what was asked for in part (a). Lots carefully showed what they were multiplying by when scaling up equations, making things clear throughout their solution. Some, however, assumed independence as evidenced by the sight of xy. Some tried to use a Venn diagram with $(0.65 - x)$ and $(0.65 - y)$ but then could not manipulate the expressions correctly in a formula.

A second equation usually found accurately in part (b)(i). Of those who did not get a second equation, they were either left with an equation with $P(B \cap C)$ still in it and did not know how to proceed having not understood that B and C were mutually exclusive.

In part (b)(ii) the correct equations were usually solved correctly and both marks scored. Of those who had an incorrect second equation, most tried to solve and gained the method mark for showing substitution or clear elimination.

When determining whether two events are independent it is essential for students to label each probability so that it is clear which test for independence is being used. The question also required students to show their working clearly so if using $P(A \cap B)$, they needed to show how it was calculated. Of those who did not achieve the marks, it tended to be because of a lack of labelling or for using incorrect notation (union instead of intersection).

Question 7

Part (a) was an accessible mark for those who understood cumulative probabilities. There were quite a few students though who tried to use the sum of the F(x) values to calculate k. Many of these abandoned the attempt and eventually used the correct F(4) =1. A small number of students attempted to verify k=4 but were not always sufficient in their conclusion.

In part (b) many were able to find the correct probability distribution for X. However, it was clear that many did not know where to begin. The most typical errors were those subtracting all the previous F(x) probabilities rather than the preceding one alone and only finding P(X = 2) correctly.. Some of the less successful students did think to substitute in the given value of k, but then incorrectly treated these as P(x) rather than F(x) for the rest of the question.

Part (c) was well attempted with 4 being stated as the mode or a correct follow through value if they had made an earlier mistake. The most common mistake was to see the probability as the answer here, rather than the x value that corresponded to the highest probability. Another error seen in this part was to calculate E(X) rather than find the mode.

The vast majority used the correct method to calculate E(X) and E(X^2) in part (d), only losing the first 2 M marks if they'd previously been unsuccessful in (b) and were using F(x) values rather than P(x). The majority were successful in their attempt to find Var(X), with the most common mistake being to fail to square the mean in the subtraction. Occasionally the final step was omitted or incorrect (for example multiplying by 13 rather than 169), but this step for the 4th M mark was generally well done. It was unusual to find students using the distribution function for Y, but some did use this alternative method.

Question 8

This proved to be the most discriminating question on the paper.

Of those who attempted part (a) many were able to gain full marks with the help of a sketch. Others appeared confused by the given statement not knowing what to do with the 0.6 probability and were not able to standardise correctly. Many incorrectly divided by the variance instead of the standard deviation. Others failed to use the full z-value from the tables but used a rounded value instead.

Most students were able to give a correct standardisation of $\frac{\frac{3}{2}\mu-\mu}{\sigma}$. From this point even with correct answers clarity and brevity were often lost. The simplest route would have been to equate this to 1.5 obtaining $\mu = 3\sigma$ and solving with the given equation $2\mu = 3\sigma^2$ to give $\frac{2\mu}{\mu} = \frac{3\sigma^2}{3\sigma} \Rightarrow \sigma = 2\ and\ \mu = 6$. However frequently complicated algebra hindered progress and caused errors. Some students sometimes incorrectly assumed $\sigma = 6$ from (a).

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