Examiners' Report
Principal Examiner Feedback

January 2023

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

January 2023
Publications Code WST01_01_ER_2301
All the material in this publication is copyright
© Pearson Education Ltd 2023

General

All questions on this paper allowed all students to make some progress, but questions 3(b), 4(ii), 5(e) and 6(d) proved to be more challenging. The questions requiring a comment or an explanation in words were often not answered very well and sometimes not even attempted. Students need to understand that a full method should be given for 'show that' questions.

Report on Individual Questions

Question 1

(a) This was a straightforward opening to the paper with the vast majority of students able to use the histogram to complete the given frequency table. The majority of those that were unsuccessful failed to realise that the area of bar represented the frequency and so simply gave frequencies of 25 and 5.

(b) Again, this part was answered well by many students with a variety of approaches seen. Those that worked with frequency densities were far more successful than those that worked with area.

(c) It was disappointing to see that some students were unsure on how to calculate a mean from a frequency table. Some added the frequencies and divided by 5 whilst others used the sum of class width multiplied by the frequencies. Those that knew exactly what to do usually scored both marks.

(d) Again, it was disappointing to see that some students could not use the correct formula for the standard deviation. The most common errors included $\sqrt{41033-\left(\frac{1891}{100}\right)^2}$ or $\sqrt{\frac{41033}{100}-\left(\frac{1891^2}{100}\right)}$. It is worth noting that whilst the mark scheme allowed use of 18.9 as the mean, students should be encouraged to work with exact answers in their working of calculations.

(e) A range of different approaches were taken to find the lower quartile. Those that worked with n rather than $n + 1$ were usually more successful. In some cases, students lost marks by stopping once they had found the lower quartile and did not go on to calculate the interquartile range. Students are advised to read the question carefully and check that they have answered the given question.

Question 2

(a) Many students were able to complete the tree diagram from the information given. Common errors seen were to repeat the $\frac{4}{8}$ (given in the question) following the 1st was green onto the branches following the 1st was blue. Also, on the second branches $\frac{5}{8}$ and $\frac{3}{8}$ were sometimes given the wrong way round. If marks were lost, then often the 3rd branches contained errors. A common error included writing the probabilities the wrong way round.

(b) The follow through in the mark scheme allowed those students with incorrect tree diagrams the opportunity to gain the method mark, which they often did. Many students were able to score full marks, with some taking the onerous route of calculating $\frac{5}{9}\times\frac{4}{8}\times\frac{7}{13}+\frac{5}{9}\times\frac{4}{8}\times\frac{6}{13}+\frac{4}{9}\times\frac{5}{8}\times\frac{7}{13}+\frac{4}{9}\times\frac{5}{8}\times\frac{6}{13}$.

(c) Again, the follow through in the mark scheme allowed those students with incorrect tree diagrams the opportunity to gain the method mark. Surprisingly, some students after calculating the two products failed to add these together.

(d) This part of the question proved to be more challenging to some students, with some students failing to realise that this was a conditional probability question. Common errors seen were $\dfrac{\frac{5}{9}\times\frac{4}{8}\times\frac{8}{13}+\frac{61}{234}}{\frac{61}{234}}$ and $\dfrac{\frac{5}{9}\times\frac{4}{8}\times\frac{8}{13}\times\frac{61}{234}}{\frac{61}{234}}$.

Question 3

(a) Students answered this part of the question well as they were able to find, in terms of a, a correct expression for E(X) and so many scored both marks.

(b) This part of the question was not answered well and students found this very challenging. Many left this blank as they failed to realise that $0 < a < 0.6$ was required. For those that did attempt this part, a common incorrect approach was to use $0 < a < 1$ which of course is a requirement for the pdf but not for E(X). Those that realised $0 < a < 0.6$ usually scored full marks.

(c) In this type of question, it would be worth encouraging students to show their working clearly. Often the expression for E(X^2) was embedded in an attempt to find Var(X) and whilst the mark scheme allowed for this, students often made arithmetical errors when taking this approach. A common error when finding the Var(X) was to not use E(X)^2 and led to $\mathrm{Var}(X)=13.2-12a-3.6-2a$ which meant that no further marks were available. Those that used a correct method sometimes made arithmetical errors in obtaining a 3-term quadratic. It is worth highlighting that a full method should be shown when solving their 3-term quadratic as many students relied entirely on using their calculators to solve the equation.

Question 4

(i) (a) Many students were able to score at least 2 of the 4 marks available in this part of the question, usually for finding $p+q=\frac{7}{25}$ and $q+r=\frac{1}{5}$. Those that found all 3 equations often lost the final mark as they simply stated the given answer or wrote no further working. As this was a 'show that' question then students should be encouraged to show all their steps in achieving the given answer.

(i) (b) This part of the question was answered well with a range of marks available to all students. Even those students that could not access part (a) were able to find some of the required probabilities. The most common probability not found was $s = \frac{3}{5}$ as this required use of the sum of the probabilities equals 1.

(ii) This part of the question proved to be challenging. The way the question was worded directed students as to what they needed to do to start the question and, in many cases, this is all they could do. As a result, the majority of students scored either 1 or 2 marks. Only the stronger students were able to access the remaining 2 marks. The students that took the route of writing the expression as $1+\frac{25}{x^2+5x}$ were usually the most successful. It was obvious that some students were confused with the requirements of mutually exclusive events and often used the properties of independence instead. Again, as this question was a 'show that' question students should be encouraged to show all their steps in their solutions.

Question 5

(a) This was the standard normal distribution question and was generally answered well by the majority of students. A significant number of students lost 2 marks as they failed to subtract from one the value obtained from the normal tables. A simple diagram would have helped many to avoid this error.

(b) (i) Many students were able to score well as they were able to standardise and set equal to a *z* value within the given range. The most common error was the use an incorrect *z* value of 0.68.
(ii) Many students were able to score this follow through mark but many wasted time and standardised again, when they could have used the symmetry of the normal distribution to find the lower quartile once they had found the upper quartile.

(c) The first mark was often scored by the vast majority of students as they were able to use their lower and upper quartiles found previously in an expression for $1.5(Q_3 - Q_1)$. Students who lost marks often failed to show sufficient working and ended up with answers outside of the given ranges. As the definitions for outliers were given students should be encouraged to show their substitution into these expressions.

(d) The demand for this question asked students to use standardisation and this is what was required for the first two marks. A significant number of students failed to do this and therefore lost all 3 marks. Those that did often scored the first two marks. As this question was a 'show that' question students should be encouraged to show all their steps in their solutions. Too many quoted the given answer without sufficient steps for the final A mark to be awarded.

(e) This part of the question proved to be more challenging for many students. Again, like question 2, a common error was that students failed to realise that a conditional probability was required. A common error was to find $\mathrm{P}(L > 5)$ and go no further. Even those students that recognised that conditional probability was required often failed to realise that greater than 5 given that it is not an outlier required the probability $\mathrm{P}(5 < L < 5.58)$. A few students recalculated $\mathrm{P}(3.42 < L < 5.58)$ even though the answer was given in part (d)

Question 6

(a) As this required an interpretation in words many students found this difficult. Too many students referred to positive correlation or as *x* increases then *y* increases. Some students mixed up the units (grams and °C) and some failed to state the required units completely.

(b) This was a straightforward part to the question, and many scored full marks. Common errors include the substitution of 80 or 100 rather than 90 as stated in the question.

(c) Again, as this required a comment using words some students found articulating their answer difficult. Those that were successful usually made reference to 90 being outside of the range and therefore unreliable. Students should be discouraged from using the word 'it' as in many cases it was not clear whether they were referring to 90 (which was required) or 396 (which was not required).

(d) This proved challenging for many students. However, there were a range of marks seen depending on students understanding of the requirements of the question. To answer this question fully students needed to the use of the regression line. To find S_{xx} students needed to use the regression line to find either $\sum x$ or $\bar{x}$. To find S_{xy} students needed to use the gradient of the regression line. Those that failed to realise either of these often scored one mark only, usually for finding S_{yy}. Those that realised both of these scored at least 6 of the 7 marks available. Again, this was a 'show that question' and students were required to show all their steps in their solution. Too many did not show their substitution into $r = \frac{S_{xy}}{\sqrt{S_{xx} \times S_{xy}}}$ and just quoted the given answer and so did not score the final A mark.

(e) This part of the question required a worded response again but was done better than the previous two parts. Many students were able to score one mark. This was usually for stating that the scatter diagram showed positive correlation or that *r* was close to 1. Rarely were both aspects given and when they were some students failed to discuss the suitability of a linear regression model.

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