Examiners' Report
Principal Examiner Feedback

Summer 2024

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 2024
Publications Code WST01_01_2406_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. If asked to use standardisation then the standardisation should be shown. 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) was generally answered well. Many students were able to state the correct value of k. Common incorrect answers included $k = 5$ and $k = 43$

Part (b) was answered well. Many students were able to find Q1 and Q3 from the given stem and leaf diagram. However, some students read the stem and leaf diagram the wrong way round and so incorrectly identified Q1 = 31 and Q3 = 51

Part (c) allowed incorrect values to be followed through so many students scored M1. A few students failed to give sufficient detail in this “show that” question to be awarded the final A mark. Usually this was because after identifying the lower and upper outlier boundary, students failed to state that 85 was the only outlier, which was required as the question asked students to show that there was only one outlier.

Part (d) Many students were able to score the 1st M mark as many drew a box with two whiskers. Whilst many students were able to score full marks, often students lost marks in this part for a variety of reasons. Some drew whiskers that ended in the wrong place, some made errors in plotting Q1, Q2 and Q3 and some failed to plot the outlier.

Part (e) caused students more challenge as the question asked for a comment on the difference between the distribution of the growths of the two trees. This question also asked students to state the values of any statistic used. Surprisingly too many students failed to give supporting figures and so lost this mark. At this level students need to realise that a question like this will require some context, in this case height, a reference to a named statistic and supporting figures.

Part (f) was generally not answered well, with only the best students scoring all 3 marks on offer. Some students were able to identify a correct range, but very few could identify both of the ranges required. As the MS required the correct inequality signs for the final mark, even those students that scored marks for both ranges often failed to give a final correct range and so the final A mark was withheld.

Question 2

Part (a) was answered well by the majority of students with many identifying the correct probability (0.45).

Part (b) was a "show that" question and students should be advised that without sufficient working shown, they will not score full marks. Some students lost marks as they failed to show that $E(X^2) = 25.3$. This was required as students could state this by working backwards from the given information. Others lost marks as they failed to write down a correct method for the variance of Var(X), often due to not subtracting mean squared.

Part (c) Many students were able to use the required information to find the values of a and b and those students scored full marks. The most common error seen was $a \times 4.14 = 66.24$ rather than $a^2 \times 4.14 = 66.24$ which of course led to an incorrect value for a and consequently an incorrect value for b.

Part (d) Students found this part of the question challenging, with only the more able students scoring full marks. Many students tried to answer the question by writing down a probability statement, without thinking through all the required combinations. A common error was to include the draw, which obviously should not have been included as the question asked for the probability that Sam wins.

Question 3

Part (a) Many students were able to identify that the width of the required bar was 1.25 and so scored B1. A variety of approaches was seen for the height and generally those students that used a frequency density approach were usually successful in identifying the correct height of the bar.

Part (b) It was pleasing to see that many students are well prepared in using linear interpolation to find a median. Many students scored full marks. A few students used $n + 1$ rather than n in their calculations, but the mark scheme allowed for this, and so again full marks were usually achieved. Marks were often lost when an incorrect method was used.

Part (c) Again, this was a "show that" question and students should be advised that without sufficient working shown they will not score full marks. Students that took the approach of showing the mean of y was 2.08 and then used this value to show that the mean of w was 40.8 were usually more successful than students who took other approaches. Very few students took the approach of finding $\sum(w-20)$ and for those that did it was with mixed success. Those students that took the approach of finding $\sum w$ often lost marks as they thought that $\sum w = 10 \times 104 + 20$ rather than $\sum w = 10 \times 104 + 50 \times 20$ and so lost both marks.

Part (d) A variety of approaches were seen in this part of the question. By far the most successful approach was realising that 10 × sd of y = sd of w. Students who took this approach usually scored full marks. Common errors included incorrect use of variance expressions and failure to realise that $\text{Var}(aX) = a^2\text{Var}(X)$

Part (e) caused students some issues. Many students were able to identify that the mean wouldn't change, but too many stated it would increase or decrease and so then lost 2 marks as the final mark was dependent on the previous 2 marks. Again, many students were able to identify that the standard deviation would decrease, but too many stated it would be unchanged or increase. Only the more able students were able to give a correct reason as to why the standard deviation would decrease.

Question 4

In general, this was the most accessible question for all students and probably the most familiar.

Part (a) was answered well with many students calculating correct values for S_{dg} and S_{gg}. A few students after writing correct expression for S_{dg} and S_{gg} then made arithmetic errors and gave inaccurate values for at least one expression. As the question asked for exact values to be given, then $S_{dg} = \frac{302628}{25}$ and $S_{gg} = \frac{45471}{5}$ were allowed but usually we would expect these to be given as decimal answers.

Part (b) was answered well with many students able to calculate a correct value of the product moment correlation coefficient. Common error included the omission of the square root in the denominator.

Part (c) Again, this was a "show that" question and students should be advised that without sufficient working shown they will not score full marks. Many students were able to show the given regression line but too often students lost the final A mark as they failed to give values to the required degree of accuracy. $b = \frac{12105.12}{16769.78} = 0.722$ was not accurate enough to gain the final mark due to it being a "show that" question. Students should be encouraged in these types of questions to give answers to at least one more decimal place than the given value.

Part (d) Generally answered well as students gave a correct interpretation of the gradient. A few lacked the context required, whilst others gave the interpretation the wrong way round and so lost this mark.

Part (e)(i) Many students gave the correct answer of 138.2. Common error included an incorrect answer of –40.495 which was from not converting 2.5 metres into centimetres.

Part (e)(ii) This proved a little bit more challenging for students. The most common correct approach seen was substituting a correct value into the regression line to get a girth = –6.2 and stated that this was impossible. Common incorrect answers included reference to extrapolation.

Part (f) This proved to be the most challenging part of the question and only the best students scored full marks. The common error included substituting a girth of 17.3 into the regression line and then solving to give a length of 82.548…

Question 5

Part (a) was answered well by the majority of students, and it was pleasing to see that students followed the instruction that the standardisation was needed. A few lost the final mark as they went on to subtract the correct answer from 1, which of course is P($X > 18$) and not P($X < 18$) which is what was required.

Part (b) was answered well by many students, and again many followed the instruction that the standardisation was needed. The most common error included the use of an inaccurate z value. Students should be reminded that when values are required from the tables, they need to be 4 decimal places. A common error was to use z value = 0.25 which scored M1B0A1.

Part (c) was the most challenging part of the question. Only the more able students scored full marks. However, many students were able to score the 1st M mark as the correct probability 0.975 was seen. Some students then went on to score the 2nd M mark as they set up a correct equation leading to a value for σ. Finding $\mathrm{P}(T > \mu - 5)$ was the sticking point for many students and a variety of incorrect attempts were seen, usually based on incorrect use of symmetry. It was not unusual for students to give up at this point. Those that did persevere occasionally scored the 4th M mark as they calculated $\frac{p}{0.975}$ with a value of p within the given range stated in the mark scheme.

Question 6

Part (a) was answered well by many students as they could correctly identify the required probability from the Venn diagram.

Part (b) For those students that understood the relevance of C and S being independent, usually set up correct probability expressions and solved to give a correct value for p. For those that did not understand the relevance of C and S being independent were limited to the first 2 M marks. Often these students struggled to give 3 correct probability expressions and it was not unusual for only M1 to be awarded.

Part (c) As this allowed follow through from part (b) many students were able to score full marks.

Part (d)(i) was answered well by many students as they could correctly identify the required probability from the Venn diagram.

Part (d)(ii) was not answered as well as a conditional probability was required, and a variety of incorrect answers were seen.

Part (e) was the most challenging part of the question and many students struggled to score any marks. Many students failed to relate 76 to $\mathrm{P}(C \cap G)$ and so failed to access the question. A few did score the second M mark as they realised that they needed to multiply the number of teenagers by 0.4. The most common error seen was relating 76 to $\mathrm{P}(C \cup G)$ rather than $\mathrm{P}(C \cap G)$ and so an incorrect answer of 55 was often seen but did score M0M1A0.

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