Examiners' Report
Principal Examiner Feedback

January 2022

Pearson Edexcel International A Level
Mathematics in Statistics S1 (WST01)
Paper : 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 candidates at: www.pearson.com/uk

January 2022
Publications Code WST01_01_2201_ER

General Introduction

There were opportunities on this paper in all the questions for all students to make some progress but questions 2 (b), 2(c), 6(e), 6(f) and 7(c) proved to be more challenging. The questions requiring a comment or explanation in words were often not answered very well and sometimes not even attempted.

Report on individual Questions

Question 1

A very disappointing number of students gave answers as frequencies rather than probabilities. The set notation was not well understood by many students and the notation used was not always clear.

(a) This was quite well answered although a common mistake was to use the value of 93 for $\mathrm{n}\left(A \cap B \cap C'\right)$ rather than 93 + 1 + 9

(b) This was very well answered with the majority of students scoring this mark.

(c) Whilst there were many correct responses to this part there was a sizable number did not include the 93 in their calculations.

(d) Once again, the omission of 93 was often seen, suggesting that students were finding the probability of one defect and not the probability of at least one defect.

(e) Students who scored both marks in (d) almost always gained full credit in this part. Those who did not score in (d) regularly gained the method mark here by correctly following through their answer to part (d) by using it in the conditional probability.

(f) The transition to this part was demanding and only the best students managed to fully grasp what was being asked, appreciating that this was a probability distribution question. Some students were able to correctly tabulate X and $\mathrm{P}(X = x)$ in order to find $\mathrm{E}(X)$. The Special Case method mark was awarded fairly often, in which students at least realised that some form of probability distribution was required.

Question 2

This question often resulted in scores of 3/6 with students coping well with a standard question and less successfully with the demands of a more unusual line of questioning.

(a) Students are very well prepared for this type of question on this topic. It was very well answered by the vast majority of students and the only cause for criticism is for a small number of responses which were not accurate enough, with 2 significant figures stated, rather than the required minimum of three figures.

Parts (b) and (c) consisted of three demands, all of which were of the form "write down ..." It would benefit students if teachers were to pay particular attention to this form of questioning, stressing to students that this always means that there is no working necessary and the response they need to make is already there in front of them, either because they have some working from a previous part that simply requires a comment or there is something in the wording of the question which highlights the quick response that will secure the mark.

(b) Part (i) was an easy mark for many although quite a few carried out copious and meaningless calculations, all to no avail. Part (ii) rarely scored the mark, indicating doubts about the student's full understanding of correlation.

(c) This was another mark which very few obtained. Many were saying that there will be no change to the coefficient and simply wrote down the same answer they had for part (a).

Question 3

This question proved accessible with the majority of students able to gain marks in the earlier parts of the question.

(a) This was well answered by many students, with the appropriate level of working shown for a 'show that' question. The data in the stem and leaf diagram made it easy to identify the correct quartiles and quite a few students also found the unnecessary median. A small number of responses did not actually state what the question had asked them to show, so they said that the outlier was greater than 138.5 but did not say that c = 9, hence losing the final mark. There were some attempts where the number 9 was created from other, incorrect deductions, for example $Q_3 - Q_1 = 9$ so c = 9.

(b) A single mark was quite common here with many only calculating the value of $\bar{x}$ with no consideration of the coding and hence not adding the 125 to find the mean number of deliveries. Another quite common response saw an incorrect value of $\sum d = 29$ being used, again with no attempt to use the given coding.

(c) It was very disappointing to see so many responses indicating a very poor grasp of the basic statistical concept of standard deviation. One key clue was that this part only carried 2 marks and so very little working would be needed and yet many students went to great lengths working backwards from the given value for $\sum (x-\bar{x})^2$ to find the value of $\sum x^2$ usually making errors in this calculation. Occasionally students omitted the square root and so had a value for variance instead of the required standard deviation.

(d) This proved to be a very challenging question, with only a small number of correct solutions being seen. Many tried to use some form of normal distribution. Very few were able to state an appropriate conditional probability statement involving either X or D with expressions often containing both of the variables.

Question 4

Most students were able to make a reasonable attempt at this question, provided they distinguished between W, X and Y correctly.

(a) This was a relative easy start to the question with most students knowing exactly what was required and were able to gain the mark for this part.

(b) A lack of quality in the presentation of solutions lead to lost marks in this part of the question. Students need to be aware of the importance of correct labelling when multiple random variables are under consideration. It was quite common for students to launch into calculations without giving any indication of what they were finding. This resulted in correct calculations for E(W) being left with no label, and as a final answer. Students who found the values for the random variable X usually went on to find the correct value for E(X) as they had noted the change to the random variable.

(c) There were various methods seen for this part of the question, with some students setting up an inequality in W and others comparing the values of X and W. Both methods were generally successful. Some students reached a stage where they had $\mathrm{P}\left(W \geqslant \frac{5}{3}\right)$ and didn't know where to proceed from there.

(d)(i) It was surprising how many students did not seem to understand what is meant by a probability distribution. It was common to see a list of values for Y with no associated probabilities. A small number of students got confused and used $\frac{1}{Y}$ as the probabilities and worked with the original values for X.

(d)(ii) Students who had answered part (d)(i) often demonstrated that they knew how to find $\mathrm{E}(Y)$ and $\mathrm{E}(Y^2)$ from a probability distribution. Even if a correct expression for variance was quoted, some students forgot to square the E(Y) after substitution, typically $\mathrm{Var}(Y) = \mathrm{E}(Y^2) - (\mathrm{E}(Y))^2$, leading to a negative answer. These students showed a lack of understanding by not realising that variance is always positive and continued with a negative variance rather than checking their working.

(e) Students who used their negative variance in part (d)(ii) in this part were not able to gain marks as only positive variances were followed through. A small number of students treated the variance in the same way as expectation and calculated 2 – 3Var (Y) However, most students seemed to be aware of the correct method and multiplied their Var(Y) by 9.

Question 5

Students were usually able to answer part (a) correctly, but struggled with the rest of the question. Some students penalised themselves by a lack of understanding what the correct labelling should be, and how important having the correct > and < are in selecting the regions.

(a) This part was accessible to most students who had some understanding of the normal distribution, and it was common for full marks to be achieved in this part. A few students did not realise that they needed to subtract the probability they had obtained from 1, but they were in the minority. Use of a quick diagram would help students to recognise whether the probability they are looking for is less than or greater than 0.5

(b) Many students failed to interpret the information given in the stem to this part correctly with many of them working with a probability of 0.16 rather than 0.4
Some students realised they should be doing something with the 0.16 but divided by 2 rather than square rooting. Another common error was to 0.16 from 1 and use 0.84 as a z-value. Students who correctly worked with 0.4 usually got the correct final answer, but some got B0 for using e.g. 0.25, rather than the more accurate answer from the percentage points table. It was also common for students to use the tables in reverse, therefore working with a probability of 0.5636 but treating it as a z-value.

(c) Students had mixed success with this part of the question. Some did not really know how to get started and a number were unable to identify that in order to find the probability of a value being negative they needed to find $P(Y < 0)$
Many students who realised that 0 needed to be used in the standardisation, usually gained the first M1 but made no further progress. Others who had been unable to find the probability understood what they needed to do with it.
While some students had the correct idea of what needed to be done students would be advised to write down an expression for what they are trying to find i.e. the equivalent of "1- P(no negative values)' otherwise they need to have a completely correct method.

Question 6

There was a mixed response to this question. Although the majority were able gain marks for part (a) only the more able were able to gain marks for the later parts.

(a) The majority of students were able to substitute 62.4 and evaluate the mean number of marks.

(b) The majority of students gave an interpretation of some sort however, many lacked detail in their answer, for example simply stating that "the higher a student scored on one test, the higher they scored on the other". Others described the relationship as "positive correlation". Only about 30% of students explained specifically what the gradient, 0.748, of the regression line represents.

(c) Many students appreciated the statement was not reliable with about 30% combining this with an appropriate reason. The most common reason given referred to extrapolation or out of range. However, a typical error was to substitute 0 in and obtain the 10 .8 and wrongly conclude that this meant the line was reliable.

(d) This part was not answered well – students were not fully aware of needing to find the point whose vertical displacement from the line of best fit was the greatest.

(e) It was quite rare to see an answer that scored any marks here. Very few spotted the need to set up the inequality $p < f$ and those that did were often unable to see how to replace f other than by substituting in a value. Quite a few students wrote down a value with no apparent reasoning.

(f) This part really enabled the most capable students to demonstrate their understanding and almost all the students who were able to obtain the first two method marks for $\sum pf$ and $\sum f$ were able to obtain the value for b accurately. A smaller number were able to get one of these two values and some students discounted the changed values altogether and simply used the original data.

Question 7

This question proved to be a good discriminator with only about 15% of students producing a fully correct solution.

(a) Many students were unable to recognise how the cumulative distribution related to the equation. One of the most common incorrect approaches was to attempt to solve the given equation. Students coming up with a product of three probabilities were almost always able to find the necessary equation to complete the proof.

(b) The students who attempted this almost always secured the mark. Those that didn't often attempted to solve the equation proved in part (a) not recognising the significance of the word verify, or recognising that less work might be expected for 1 mark.

(c) Students who understood what was meant by a cumulative distribution were often able to score the first couple of marks available for finding the value of a. After that the most common error was failing to deal with the number of possible arrangements in the calculation. A minority of those students who managed to find $P(X = 1)$ or $P(X = 2)$ left these as their final answer showing a lack of understanding of the relationship between $F(x)$ and $P(x)$.

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