Pearson Edexcel

Examiners' Report
Principal Examiner Feedback

January 2019

Pearson Edexcel International Advanced Level
In Statistics S1 (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 students at: www.pearson.com/uk

Grade Boundaries

Grade boundaries for all papers can be found on the website at:
https://qualifications.pearson.com/en/support/support-topics/results-certification/grade-boundaries.html

January 2019
Publications Code WST01_01_1901_ER

General introduction

The paper proved accessible with opportunities for all students to score in each question. Questions 3(b), 4(f), 5(ii) and parts of 6 proved quite discriminating whereas question 2 was the most straightforward question on the paper with 63% scoring full marks.

Report on Individual Question

Question 1

Many students could identify a pair of mutually exclusive events in part (a) but some lost the mark for writing P(B) and P(C) or P($B \cap C$) = 0. The rule for independence was well known and usually applied correctly in part (b) to find p and the follow through marks in part (c) enabled those who had made a slight slip to pick up marks here too. It is disappointing to see some students giving negative values for q, r or s somehow losing sight of the fact that these values are probabilities. Part (d) proved to be quite challenging and many did not realise that a conditional probability was required, possibly because we did not use the phrase "given that" explicitly in the stem this time. Those who did identify the correct conditional probability usually had a correct denominator but sometimes one of the probabilities on the numerator was missing. A number of students lost the final mark for giving their answer as 0.581 rather than the exact fraction as required.

Question 2

This question was very accessible and many students scored well here. The usual problems with minus signs caught a few out but the vast majority clearly understood the topics tested here. Part (a) caused few problems though sign slips for some meant that the as did not cancel out and even those who did not manage part (a) were often able to answer (b) correctly. In part (c) those who used the Var(Y) = (– 2)^{2}Var(X) were usually able to obtain the correct answer but those who found the distribution of Y and tried to find Var(Y) directly rarely gained full marks.

Question 3

There were still a number of students who do not handle the notation connected with a normal distribution correctly (for example equating probabilities to z values) but we saw a good number of students achieving correct answers to part (a) and often part (c) as well. In part (a) the standardisation was usually correct but a few failed to subtract the probability from the tables from 1. To make progress in part (b) we required a correct conditional probability statement. Several simply found P(W < 49) or P(49 < W < 51) without identifying a conditional probability. Others tried to find P(W < 49 | W < 51) but a number did find the correct ratio. Some of these used the tables correctly but their calculators "simplified" the answer to the exact fraction 5/12. Students should have been aware that such an answer involving a ratio of normal probabilities would not be an exact fraction and a mark was lost if they did not give their final answer as a suitable decimal. There were many correct answers to part (c) but some failed to use the percentage points table (or show that they worked to an equivalent degree of accuracy on their calculators) and had an equation of the form $\frac{H-64}{8} = 1.28$ which would have lost the B1 mark. Some students mis-read the table using the 1% point instead of the 10% point and others used the correct z value with an incorrect sign failing to realise the inconsistency of having a minimum weight to be in the top 10% that was <u>below</u> the mean. A diagram may have helped them spot this.

Question 4

In part (a) most students were able to identify that the width was simply 4 times 0.5 cm but the calculation of the correct height still causes problems and many made quite complicated calculations before arriving at their (only occasionally correct) answer. Probably the most successful approach was to realise that if the area for the given frequency of 25 was 3 cm^2 then the area for a frequency of 24 would be 0.96 of this or 2.88 cm^2 and hence the height must be 1.44 cm. In part (b) the linear interpolation was often correct but errors did occur sometimes with class boundaries or cumulative frequencies. The mean was a straightforward calculation in part (c) and it was encouraging to see a good number also calculating the standard deviation correctly too in part (d) though some are still using an incorrect formula such as $\sqrt{\sum \mathrm{f}x^2 - \bar{x}^2}$. In part (e) most chose the simple route of comparing their answers from parts (b) and (c) but some engaged in a lot of extra work to find Q_1 and Q_3 and make a comparison using $Q_3 - Q_2$ and $Q_2 - Q_1$. In part (f) many candidates did not attempt to give any reasons to support their statements and some did not identify the direction of any changes. The key feature to mention in relation to the median was that the 18 adults we were given extra information about were below the median beforehand and would remain below the median so there would be no change. Many did mention that the effect on the mean would be to lower the overall total and so reduce the mean and in connection with the standard deviation we simply wanted a realisation that these 18 values were moving further away from the mean and so spread, or the standard deviation, would increase.

Question 5

This question was the most challenging with 30% of the students scoring 1 or fewer marks. In part (a) most obtained the correct value but a small minority simply gave the answer as 1.5 from $\frac{0+1+2+3}{4}$

Those who completed the table in part (c) usually scored the mark for realising that the sum of the probabilities must equal 1 but here were a number of slips in part (b) with 0.4^2 or $0.4 \times 0.4 \times 0.6$ common incorrect answers. In part (d) a good number could identify the correct values of the p_i and this often led to a correct demonstration for (i). Part (ii) proved more challenging and a common incorrect answer was $1 - p_3$. As this was a "show that" question we needed to see the 3 cases and their associated probabilities clearly identified. Many did not attempt part (e) but those who had successfully answered part (d) were often able to find $\mathrm{P}(T = 0)$ quite easily and hence complete the distribution. Part (f) proved quite challenging and few proceeded in the simplest manner by calculating relative frequencies from the data at the start to compare with the two probability distributions. A few did score a mark under the special case ruling for explaining that Ting's model was more appropriate because it included the idea that the probabilities increased as the number of throws increased but the question was really looking for a comparison of the probabilities or expected frequencies based on the models with the observed figures given at the start to provide evidence to support their choice of model.

Question 6

Part (a) simply required a comment about the means and the standard deviations of the two sets of marks. We would usually expect correct terminology to be used though for the first mark we allowed a comment that on average marks on paper 1 were higher than on paper 2. Most students gave acceptable explanations or suggestions in part (b). The calculations in part (c) were usually correct though a small minority believed that the mean of y would stay the same. It was surprising how many students went to the scatter diagram and wrote down all the points and then used these values to find the new means rather than simply use the given means in the question. In (d) (i) most gave a suitable explanation based on the fact that the xy value at (38, 0) would be 38×0 and so no change to the value of $\sum xy$ but in (d) (ii) many students used a rounded value for their new $\overline{y}$ and therefore did not show that the value of $S_{xy}=1169.8$ exactly. Parts (e) and (f) were familiar territory for most students and these parts were usually answered very well. A small number still gave their answer to (e) to only 2sf and lost a mark and there were often accuracy slips in part (f) that meant the final accuracy mark was lost. In part (g) we were looking for a realisation from students that a value of the correlation coefficient close to 1 means that the points lie close to a straight line and therefore the increase in this value would show that the points lie closer to the straight line. Few mentioned this important feature which would support Chetna's decision to omit the point before calculating the regression line. In part (h) most students who were still engaged with the question were able to secure the mark for using the value of x = 38 in their equation.

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