Examiners' Report
Principal Examiner Feedback

October 2023

Pearson Edexcel International Advanced Level
In Statistics (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

October 2023
Question Paper Log Number 74325
Publications Code WST01_01_pef_20240118

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".

Report on Individual Questions

Question 1

Part (a) provided an accessible start to the paper with many students scoring full marks. Those that did make errors usually made errors when completing the third branches of the tree diagram. It was surprising to see that some students showed a lack of appreciation that all sets of branches should sum to 1.

Part (b) was answered well as the required information was given in the tree diagram.

Part (c) required using branches filled in by students and so caused a few more issues for some students. However, many students that had incorrect values on their tree diagram were able to follow the correct branches and write a correct expression for two games in a turn.

Part (d) and part (e) required use of a conditional probability and in part (d) it was pleasing to see that students seemed more comfortable with this concept. This was probably due to the fact that the parts needed to answer the question were found in parts (b) and (c). It was surprising to see that some students showed a lack of appreciation that probabilities cannot be greater than 1. On the other hand, part (e) caused students more difficulty and highlighted that for some conditional probability is still a concept that they find difficult. This was probably due to the fact that they needed to use the tree diagram to calculate the numerator of the required ratio of probabilities. The common error seen was using part (b) as the numerator of the ratio of probabilities.

Question 2

Part (a)(i) was answered well with virtually all students giving a correct value of the median. Part (a) (ii) was generally answered well. The usual error came when finding Q_3. Common error was to give Q_3 as 62 or 62.5.

As part (b) allowed for follow through from part (a) the vast majority of students were able to substitute their quartiles into the given outlier expressions. It was surprising however that a number of students who had correctly identified the outlier boundaries then failed to identify the 2 outliers that the questions asked for.

As part (c) allowed for follow through from parts (a) and (b) the vast majority of students were able to score marks in this part of the question. Errors usually occurred when plotting the ends of the whiskers or not plotting the 2 outliers, even when they were correctly identified in part (b)

Part (d) caused the most difficulty for students. Too many just quoted figures without making any comparisons. Some were able to either give a correct comparison of the median and/or the range/IQR but then could not give a correct interpretation of either of these. Some were able to give a correct interpretation of the average and/or spread but missed that the question asked for a comparison of the distribution.

Question 3

In (i) (a) attempts at completing the given Venn diagram were varied. Whilst a number of students could complete this correctly, some found finding x and y difficult. The common error for y was 0.3 which was the probability that Bob did not buy onions and this was given in the question. It then followed that students who gave y as 03 then gave x as 0.35 which was also incorrect.

The follow through in (i) (b) allowed for the majority of students to gain this mark.

As (i) (c) was a show that question students need to be encouraged to show sufficient working to ensure that all marks are achieved. Those that realised this often scored both marks but too many lacked the detail required to score both marks. A few students showed a lack of understanding of what statistical independence requires and therefore failed to attempt this part of the question.

(ii) (a) was answered well and often students gained the B mark available here.

However (ii) was not answered well and caused students the most difficulty. Those that understood that the use of the addition formulae and that $\mathrm{P}(F \cap G) = \mathrm{P}(F) \times \mathrm{P}(G)$ due to F and G being independent scored well, but too many students did not or had a miss understanding of independence. The common errors seen included using $\mathrm{P}(F \cap G) = 0$ or $\mathrm{P}(F \cap G) = \mathrm{P}(F) + \mathrm{P}(G)$.

Due to the follow through in (ii) (c) many students scored the B mark available here as they realised that $\mathrm{P}(F \cap G) = \frac{2}{7} \times \mathrm{P}(G)$, even those that failed to realise this was needed in (ii) (b).

Question 4

In part (a) many students were able to give the correct expression required for $\mathrm{E}\left(\frac{1}{x}\right)$ and so gained B1. The common error seen here was to write the expression for E(X) rather than the expression for $\mathrm{E}\left(\frac{1}{x}\right)$.

In part (b) those students that knew how to find a variance went on to gain full marks. However, some students found $\mathrm{E}\left(\left(\frac{1}{x}\right)^2\right)=\frac{5}{24}$ and stopped there as they failed to realise that $\mathrm{Var}\left(\frac{1}{x}\right)$ required the subtraction of the square of the given value for $\mathrm{E}\left(\frac{1}{x}\right)$. Some students calculated $\mathrm{Var}(x)$ in the hope that they may gain some credit here. A few students incorrectly found $\mathrm{E}\left(\left(\frac{1}{X}\right)^2\right)$ by the following incorrect method $1\times\left(\frac{1}{10}\right)^2+\frac{1}{2}\times\left(\frac{1}{5}\right)^2+\frac{1}{3}\times\left(\frac{3}{10}\right)^2+\frac{1}{4}\times\left(\frac{2}{5}\right)^2$ but did gain the second M mark as they evaluated this and subtracted the given $\frac{2}{5}$ squared.

Two different approaches were seen in part (c), one route was to use expectational algebra and the other was to find the 4 y values required to answer the question. Those that took the first approach were generally more successful and required far less work. Part (c) (i) was the easier part of the question and many students scored the B mark available here. In (c) (ii) those students that used expectational algebra generally scored both marks but those that used the 4 y values approach often failed to give a correct expression for Var(Y), often finding $\mathrm{E(Y)}^2$ only.

Part (d) caused the most problem for students. Not only did it require use of conditional probability, the P($Y < 20$) also needed to be found. Many students failed to find the P($Y < 20$) and so made no progress with this question. Those that realised that $\frac{30}{X}<20$ often incorrectly solved this inequality and wrote $X<1.5$. Some students correctly found the P(Y < 20) but then could not identify the correct conditional probability required. The most common error seen was to find $\frac{\mathrm{P}(X<3)}{\mathrm{P}(X>1.5)}$. Only the most able students scored all 4 marks available here. Those that found all 4 y values often scored the first M mark only.

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.

In part (b) many students were able to standardise using 190 and then subtracted from the given answer from part (a) to gain full marks. However too many students subtracted the correct probability for 190 from 1 before subtracting from the given answer in (a) and therefore gave an incorrect answer of 0.0968.

In part (c), those students that realised that they could use either tail and set equal to a compatible *z* value usually scored full marks. Common errors included either using an incompatible *z* value or trying to use both tails and set equal to 1.6449

Part (d) was answer well by many students. Common errors included either the use of an incompatible *z* value or using a less accurate compatible *z* value (usually –1.04)

Part (e) produced a wide variety of solutions. Better students that used compatible accurate *z* values usually scored full marks, but on occasions gave $\sigma = 14.7$ and so lost the final A mark. Again, as with other parts in this question, incompatible *z* values or less accurate *z* values were seen. It would be worth pointing out to students that accurate z values are required when answering these types of questions.

Question 6

In part (a) (i) many students were able to find the correct point of intersection of the two lines, but far too many students gave their answers as decimals rather than exact fractions, which then meant that in part (a) (ii) they could not show that $\sum x = 25$. Often 24.96 was seen which came from multiplying a decimal answer (2.08) by 12

In part (b) many students knew the correct expression for S_{xy}, but again, for those students that gave the point of intersection as decimal values inaccurate answers followed. Often $\sum y = 53.04$ was seen which came from multiply a decimal answer (4.42) by 12. Students then were able to substitute 25 and 53.04 into the correct expression so M1 was awarded, but as this gave awrt 5.52 and not 5.6 so the final A mark was withheld.

Part (c) proved to be problematic for students and only the better students were able to answer this part of the question. Some students used the gradient to find S_{xx} but then did not realise that they could use the gradient to find S_{yy} and so no further progress was made. For those students that attempted to find both S_{xx} and S_{yy} the next M mark was usually scored, as they correctly substituted these values into the correct expression for the product moment correlation coefficient.

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