Examiners' Report
Principal Examiner Feedback

October 2024

Pearson Edexcel International Advanced Level
In Statistics (WST01) Paper 01

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. Questions that state 'Solutions relying entirely on calculator technology are not acceptable' require students to show their method and not just state values from their calculator. These are usually used when standardisation is required or when solving simultaneous equations. 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) was generally answered well. Many students were able to state the correct value of the mode for action films. A few students left this blank or failed to interpret the key and gave an incorrect answer of 7. A few students gave the mode of the comedy films rather than action films.

Part (b) was generally answered well. Many students were able to find the values of a, b and c. However, some students made an error in finding one or two of these value. A few students gave the three numbers required but failed to identify which was a, which was b and which was c.

Part (c)(i) was answered well, and many students were able to find the mean for action films. Whilst the mark scheme allowed for answers which rounded to 130, students should make note of the instruction given in the question which asked students to give their answer to one decimal place. A few students ignored the fact that $\sum x$ was given and in these cases some students made an error in calculating the $\sum x$ and therefore lost this mark.

Part (c)(ii) was generally answered well, and many were able to find the standard deviation for action films. A few students lost marks as they calculated the variance rather than the standard deviation. A common error was to use their rounded answer for the mean which led to an incorrect answer of 12.9 and so lost the final A mark. Students should be encouraged to use exact values rather than rounded values in these calculations.

Part (d) Many students were able to substitute their values of the mean, mode and standard deviation into the given expression. A few students made an error in the substitution of these values and often the median rather than mode was used. The mark scheme allowed for follow through their expression for the interpretation and many gave a correct interpretation. However, some students interpreted this value incorrectly. A few students failed to evaluate the given expression and wanted to compare the mean and median in order to describe the skewness. As the question asked for the expression to be evaluated no credit was given to this approach.

Part (e) caused students more challenge as the question asked for a comment on the difference between the distribution of the running times of these two types of film. 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 'running time/length', a reference to a named statistic and supporting figures.

Question 2

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

Part (a)(i) was answered well with many students showing a correct numerical expression for S_{yy}. As the question was a 'show that' question students needed to show a numerical expression for S_{yy}. So those students that quoted that $S_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 734$ were unable to score this mark.

Part (a)(ii) 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. A few students lost the final mark as they failed to give the answer to the required degree of accuracy.

Part (b) proved a little more challenging for students. Many students failed to interpret the value of the product moment correlation coefficient found in part (a)(ii). Often students made reference to the product moment correlation coefficient being close to 1 or that this value showed strong positive correlation without actually interpreting what this meant.

Part (c) like part (b) proved a little more challenging for students. Many failed to comment on whether or not the product moment correlation coefficient was consistent with the biologist's belief. Some students interpreted the value of the product moment correlation coefficient, obviously confusing what part (b) and part (c) were asking.

Part (d) Many students were able to find the equation of the regression line. However, some students calculated an incorrect value for the gradient. Common error was to use $\frac{S_{ww}}{S_{yy}}$ rather than $\frac{S_{yw}}{S_{yy}}$ to calculate the value of *b*. Whilst the mark scheme condoned answers which rounded to 0.3, students should be encouraged to give answers correct to 3 significant figures. Some students found finding the value of *a* more challenging. Common error was to mix up $\bar{w}$ and $\bar{y}$. A few students lost the final A marks as they failed to give the final answer to required degree of accuracy.

Part (e) As the mark scheme allowed for follow through an incorrect regression line, then this part of the question was answered well. It was surprising that when using an incorrect regression line students did not question their answers. Some had negative answers, and some had answers which would suggest that the weight of the rabbit was the same as an adult. Students at this point should have realised that they had gone wrong somewhere in this question.

Question 3

Part (a) was answered well with many students scoring full marks. Occasionally numerical errors were seen. Common error was to leave the intersection of all 3 magazines blank. A few students gave the entries in the Venn diagram as probabilities, either as fractions or decimals. Whilst the mark scheme condoned this, students should be encouraged to use the frequencies rather than probabilities.

Part (b)(i) As the mark scheme allow follow through from the Venn diagram, many students scored this mark.

Part (b)(ii) As the mark scheme allow follow through from the Venn diagram, many students scored M1. Even those students who had incorrect Venn diagrams were able to use the given information to gain the correct answer.

Part (c) was not answered as well as a conditional probability was required, a variety of incorrect answers were seen. A common incorrect answer included $\frac{\frac{25}{80}}{\frac{80}{200}} = \frac{25}{32}$

Question 4

This question stated that 'Solutions relying entirely on calculator technology are not acceptable' and so standardisation was needed to be seen.

Part (a) Generally students did show their standardisation and M1 was awarded. Many went on to give a correct answer and scored A1. Common error was to give an answer of 0.8944, which came from the fact that they thought they needed to find P($X < 190$) or they forgot to subtract from 1.

Part (b) was more problematic for students and whilst many scored full marks a variety of z values were used, causing students to lose the A marks. Students should be reminded that when values are required from the tables, they need to be 4 decimal places. Common error was to use a z value = –1.28/–1.282. Some students who used a z value to 4 decimal places lost marks as their z value was not compatible to their standardisation. Common error was $\frac{d-170}{16} = 1.2816$

Question 5

Part (a) caused some students problems as they could not explain why a histogram was appropriate to represent the data. A variety of incorrect answers were seen.

Part (b) was answered well by many students. Many were able to establish a ratio between photographers and area. A few used frequency density. Those students that did not get a correct answer after establishing a correct ratio often did so as they made a numerical error in their calculations.

Part (c) 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 (d) This part of the question caused some students problems. Often, students failed to make a comparison between the mean and median or failed to make a correct compatible comment about Charlie's decision. Common errors usually included comments made about the shape of the histogram. Students should be encouraged to use the information given in the question when answering these types of questions.

Question 6

Part (a) Students fell into 2 camps here. Those that knew what a cumulative distribution function was and those that did not. Those that did found a correct value of k. Those that did not often treated this as a probability distribution and added the probabilities and thought this equalled 1.

Part (b) A similar scenario. Those that knew the difference between a cumulative distribution and a probability distribution and those that did not. Those that did often scored full marks. Those that did not often scored no marks as they simply substituted their value of k into the cumulative distribution function. Some students seemed to think that a probability distribution was E(X) and this calculation was seen.

Part (c) was generally answered well. Many followed the instructions given and formed two equations in a and b and then wet on to solve these equations to obtain the required answer. However, some did not. This was a 'show that' question so students should be encouraged to show all their steps in their solution. Also, the question stated 'Solutions relying on calculator technology are not acceptable' which meant that students needed to show how they solved the simultaneous equations. Some failed to do this and so last the final A mark.

Part (d) This was a 'show that' question and many students showed sufficient working to gain full marks.

Part (e) Many students were able to find Var(Y) using the given information. Those that did not use the given information and tried to calculate values for $5 - 2Y$ often did so unsuccessfully. Once a value for Var(Y) was found many students knew that this needed to be multiplied by 4. Common errors included finding $5 - 2\text{Var}(Y)$ or $-2\text{Var}(Y)$

Part (f) caused students the most problems. Only the best students realised they needed to use $\text{P}(X = 1) \times \text{P}(Y = 2) + \text{P}(X = 2) \times \text{P}(Y = 1)$. Common error seen was $0.1 \times 0.15 + 0.2 \times 0.15 = 0.045$ which comes from using the cumulative distribution function rather than the probability distribution function.

Question 7

Part (a) Many students were able to complete the tree diagram for the given information. Some students failed to note that the question said that the counters were selected without replacement and so filled out the 2nd branches incorrectly. A few failed to give the tree diagram as probabilities and n and $n + 1$ was seen too often.

Part (b) This question was a 'show that' question and so at least one line of working between M1 being awarded and the given answer was required and no incorrect was allowed. Quite

often the M1 was awarded but then either no further working or incorrect working was seen before the given answer, so it was not usual for the final mark to be withheld.

Part (c) was answered well with many students scoring at least 1 mark. Many were able to solve the required equation to find $n = 24$. However, some students did not go on to answer the question and give the total number of counters in the box before any were selected.

Part (d) As this question involved a conditional probability students found this difficult. The question asked for working to be shown so a correct answer with no working scored no marks. Common error was to use an incorrect numerator.

Question 8

Part (a) A variety of marks were awarded in this part of the question. Most students were able to standardise using μ and σ and setting equal to a z value within the given range. When the first A1 was awarded, it was usually for standardising and setting equal to 1.04. As the mark scheme allow a correct 2nd equation equal to a z value correct to 2 decimal places many students were awarded this mark. Those that did not usually had an incompatible z value. Students should be encouraged to use z values that are 4 decimal places. Many students were able to solve simultaneously but again students need to be encouraged to show their working. A few students found one correct value, but the second value was not given to the required accuracy so the final A mark was withheld.

Part (b) This proved to be challenging for many students. Those students that used the given information found Q1 correctly and scored B1. However, those that ignored the given information and tried to calculate Q1 often lost this mark. Common error for Q1 was 208.25/208.3. Once Q1 was found many students were able to use the given information to calculate an outlier limit. At this point many students made no further progress. Those that did usually standardised using their limit, 215 and 10. A common incorrect method was to standardise using 221.74. Quite often the final M mark was not awarded as students did not realise that due to symmetry then all that was needed was to multiply their probability by 2.