# Examiners' Report
# Principal Examiner Feedback

## June 2023

Pearson Edexcel GCE
In Statistics (9ST0)
Paper 01: Data & Probability

**General Comments**

This paper was accessible to all candidates.

**Question 1**

This question served as a straightforward introduction to the paper, with at least one of the three methods accessible to all candidates. (d) required candidates to refer to both minimising expense and minimising inconvenience to customers. Candidates generally lost marks for either only referring to one of these, or failing to realise there was a cost associated to a staff member waiting by a door. For questions such as (e) it is important to make sure that candidates are answering the exact question asked. For example, in this case a large sample of people to complete the survey would result in a large sample and more responses, but this does not impact response rate.

**Question 2**

This question was well attempted by candidates, and in general they were able to perform well. In (c) many candidates did not explain how colour would make the graph easier to understand, such as by giving each country a different colour, and merely restated that it would make the graph easier to understand.

(d) was well accessed with candidates able to give a range of criticisms. Candidates are reminded that in questions where they are required to give a number of criticisms, such as (d), they are advised to use bullet points, and to be careful not to repeat the same point with different wording.

**Question 3**

This was a relatively standard Venn diagram question. Candidates are reminded to show full working. For example, if a candidate has filled out the diagram incorrectly in (a) they still received method marks in (c) for showing full working including the use of formula, but not for simply writing down an answer with the values from their diagram.

(d) and (e) were answered well, with candidates who lost marks in (d) generally doing so for not writing a conclusion referring to the independence of the events.

**Question 4**

(a) and (b) were well answered in general, although some candidates mistakenly gave the answer a zombie would spawn in the next 10 seconds instead for (b)(i).

Candidates had a choice of the Poisson or exponential distribution for (c)(i) and (c)(ii) with most choosing to use the exponential distribution. Relatively few candidates answered (c)(ii) correctly.

(c)(iii) was more accessible to some than the preceding parts, but some candidates who had not been able to access (c)(ii) chose not to continue with the question. These candidates missed out on some relatively straightforward method marks in (c)(iii)

(d) and (e) were good at discriminating between candidates due to the difficulty of the question. Candidates were expected to use trial and improvement, as mentioned in the notes, rather than to use logarithms although this was a valid method.

## Question 5

(a) and (b) were answered well in spite of the fact that there was a misprint, though appropriate steps were taken to ensure that candidates were not disadvantaged due to this.

(c) was a straightforward question for most candidates, provided they did not get confused between the product moment correlation coefficient and Spearman's correlation coefficient.

In (d) most candidates were able to identify the fact there was a difference in the correlation coefficients, but few fully explained that this was particularly surprising as it was exactly the same data.

A full answer to (e) was an indicator of exceptional understanding and rare to come across, however some candidates did give sophisticated answers, linking together multiple ideas such as the ruralness of an area and the quality of education.

## Question 6

Most candidates filled in the tree diagram correctly, and were therefore able to access (b). The tree diagram was used by many candidates to complete (c) rather than using Bayes theorem, and as in question 3 candidates are advised to show full working in calculation based questions to earn method marks if they made a mistake in their diagram.

In (d) many candidates pointed out that the probability blue asbestos was removed was extremely high, but this is not a contextual reason – rather, this is a consequence of the reason it is removed more frequently.

(e) was generally answered well, but candidates need to make sure to link back to the statement you are commenting on in the question. Some candidates did not include numerical justification as they were instructed to and were therefore penalised. Other candidates did not refer back to safety equipment and also lost a mark for this.

## Question 7

Most candidates were able to correctly find the mean was 0.971, though some incorrectly justified this by averaging the sextiles and thus only scored one mark.

Explanations in (b) varied in quality, with some candidates manipulating numbers to attempt to find the stated value. Most correct answers considered $\mu \pm \sigma$ but there were other correct alternatives possible, by considering the z-score of a given sextile.

(c) was a simple calculation that some candidates missed as they had given up after (b)

(d) required combining two normal distributions. If candidates realised this they were generally successful as this was a straightforward example of combining two normal distributions.

**Summary**

Based on their performance on this paper, candidates should be advised to:

- give explanation answers within the context of the question
- show full working on questions requiring calculations
- use bullet points when answering questions requesting multiple reasons
- ensure they are answering the exact question asked
- attempt the whole question, even if they have not managed earlier parts