



Pearson

GQ Principal Examiner's Report

Summer 2025

Pearson Edexcel GCE

In Statistics (9ST0)

Paper 1 Data and Probability

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

Summer 2025

Publications Code 9ST0_01_2506_ER

All the material in this publication is copyright

© Pearson Education Ltd 2025

General Comments

This paper was accessible to most learners, particularly those questions that were numerical in nature. Learners often struggled however with some of the explanations that were required as they needed to be applicable to the context of the question, rather than just generic phrases and criteria that had been learnt.

Report on individual questions

Question 1

In part (a) learners were generally able to find the median and the interquartile range from the stem and leaf diagram.

Part (b) was also generally well answered, and learners were able to state the median and the interquartile range (or range) correctly. This was acceptable within the context of their explanations, but the correct values needed to be seen to gain these marks. Learners were still able to gain marks for explanation however, even if their values were incorrect as long as their values were correctly interpreted. (A few learners calculated the mean because they saw the word 'average' and then did not gain the marks as they were comparing the median of one and the mean of another).

Question 2

This question was not generally well answered for the full five marks. Learners, on the most part, calculated the mean correctly but put down the s_x value as opposed to the σ_x value. A few learners considered it as a sample (which it was not) and therefore calculated s_x , which is the sample deviation rather than the standard deviation which was asked for. When comparing their means and standard deviations if they only looked at a difference between the respective values rather than the proportions, they did not gain full marks. The majority of candidates gained 3/5 marks in this question.

Question 3

Part (a) being more numeric in nature was generally well answered by learners, although a number of learners did not gain full marks on what was essentially a straightforward probability question. The mistake that occurred in this part of the question was that learners only considered two walking and not that the third person needed to be 'not walking' so they only multiplied two probabilities together. The majority did get the second mark in the question for all three walking. Only a few learners did the alternative method of subtracting none or 1 walking from 1.

Part (b) was also generally well answered with many learners gaining full marks. Those who did not get the correct value for (i) were still able to get the final mark in (ii) if they followed through their previous value, which many did.

Question 4

This was a question where many learners did not gain marks if they did not answer the question in context. It was clear that they had learnt the generic criteria for what the properties of these distributions were but were not able to apply them to the question.

Variable A. Learners were given credit for a 'fixed number of trials' but if they commented on only two outcomes, we needed to know what these outcomes were, also the statement of fixed probability (of success) was not enough as the germination of the seed needed to be referenced for this. Many learners

were able to state the correct distribution, but it was dependent on gaining the first E mark in the question. The second E mark was for the seeds (germinating) independently of each other.

Variable B. Learners did not answer this part well. They needed to reference that as weight is continuous and Binomial and Poisson are **both** discrete distributions that neither would be suitable.

Variable C. This was similar to variable A in that learners were not answering in context but rather just reciting the generic properties of a Poisson distribution. They needed to state that there was a constant rate of cars and that they could only arrive (at the booth) singly, to gain the first mark. Many learners were able to state the correct distribution but did not gain the final mark as it was dependent on the first E mark in this part. To gain the second E mark in this part, learners needed to state that the cars arrived independently. Many learners clearly knew that this was a consideration but then explained why the cars were not necessarily independent to each other and did not gain this mark.

Question 5

Part (a) was generally well answered however several learners did not state that it was a scatter graph. Most were however able to gain the second mark for explaining that it would make it easy to see relationships/associations/trend. Those that stated correlation generally realised that scatter graph was the correct answer and got 2/2.

In part (b) learners again were able to calculate the value correctly and state that it was a strong positive (correlation).

Part (c) presented a challenge to many with a majority of learners getting marks for the special case scenario. This seems to be mainly due to learners using their calculators to get the regression line and not looking at the format required. The E mark was still available but many had the relationship the wrong way round and therefore stated that for every £1 that the house increased, the earnings increased and they did not gain this mark. They also did not gain this mark if they used the letters P and E without explanation of what these letters represented in context.

Part (d) was generally well done with most learners realising that Wales was not in the data set. Extrapolation on its own was not sufficient for this mark however.

Part (e) proved challenging for many candidate learners and they often quoted that correlation does not mean causation but could not put this into context. Some were able to state that perhaps house prices were pushing up salaries and some were able to mention inflation increasing both or perhaps employment, schools in the areas or transport hubs affecting house prices and salaries but it was a minority of learners that gained both marks in this part.

Question 6

Part (a) marks were not awarded due to lack of context. Learners needed to reference that it is the green sweets that have a constant probability, only two outcomes – green or not green, the colours of the sweets are independent to each other. General statements with no context of the colours did not gain marks.

Part (b)(i) was well answered as was (b)(ii) but (b)(iii) was not as successful as a number of learners did not realise that the probabilities for less than or equal to 19 and 9 needed to be subtracted. Learners on the whole did well on these numerical questions.

In part (c) many learners were able to calculate the mean and stated that it was approximately the same as the given number but they did not calculate the standard deviation and therefore were unable to get the final two marks in this part. Learners need to provide a complete comparison of both in order to make a

comment on Jaymini's conclusion. Only a small number of learners stated that more evidence would need to be gathered however, with most just stating that Jaymini was correct.

Question 7

Part (a)(i) was well attempted with most learners recognising that a random selection would prevent bias. Therefore avoiding the allocation of similar infants to the same group.

Part (a)(ii) learners were not awarded marks if they referred to the infant not knowing what they were being given (as the infant would not know anyway) and it was the parent or the person administering along with the researcher that would not know. A number of learners did not state that this would ensure equal treatment or equivalent statement and so did not gain the second mark.

Part (a)(iii) proved challenging for many learners. It was clear that a large number of learners did not state that in comparing the results of the two, the only difference with the two treatments was the type of oil applied and therefore a direct comparison could be made. (A number of candidates referred to the placebo as a control group - which was not specifically penalised but should be discouraged)

Part (b) learners needed to state that we needed a baseline for values (or a starting point) then a half way point (to record or withdraw if there were adverse affects eg allergies) and then the final point was not just to record and compare but to actually draw conclusions from the results.

Part (c) learners needed to compare the differences in the readings for at least two of the categories and draw conclusions. Students needed to explain or state that the paraffin and the olive oil both worked but that the olive oil was more effective. Many learners just stated that the olive oil worked and this was not enough without a comparison to the placebo results. Generally however learners gained either 2 or 3 marks in this part of the question.

Question 8

Part (a) most successful learners drew a tree diagram in order to demonstrate what their probabilities represented and this meant that they did not need to be explicit with regard to writing a given formula to gain the first marks. However the majority struggled to convert 0.1% into a decimal and although able to gain method marks for Bayes Theorem they did not get the final A mark.

In part (b) learners seemed to be confused with regard to the wording of the question and many stated that the probability of testing positive if they have the disease was low, rather than the probability of having the disease following a positive result being low. A number correctly referenced 'false positives' being large which was acceptable. It was possible to achieve full marks in (b) if it followed through correctly from their previous answer in (a)

Question 9

Part (a) was generally well attempted but if they referenced that it was symmetrical, then they needed to specify that it was symmetrical about the mean or we would accept that it was a bell shaped curve but we needed more unique and specific feature of this distribution to gain the second mark in this part.

Part (b) was generally correct

Part (c)(i) A fairly large number of learners were able to recognise that they needed to subtract the means and add the variances and that it was a normal distribution and then they were able to calculate the probability correctly. A few learners added the means and used $P(X > 50)$ with a mean of 45 which was also acceptable.

Part (c)(ii) was challenging to many with only a handful of learners achieving full marks.

Summary

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

- Give all explanations in context.
- Understand the specific properties of the Binomial, Poisson and Normal Distributions, rather than just recite them.
- Keep their working to more than 3 significant figures accuracy throughout and only rounding after their final calculation.
- Learners need to be aware of the word ‘average’ not just applying to the mean. This caused some issues in Question1 where they calculated the mean and then compared it with the median for the box plot.
- Learners should feel comfortable using tree diagrams (when not specifically requested) for conditional probability.