



Examiners' Report

Principal Examiner Feedback

Summer 2019

Pearson Edexcel GCE Mathematics

In Statistics

Paper 1: Data and Probability (9ST0_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

Summer 2019

Publications Code 9ST01_01_1906_ER

All the material in this publication is copyright

© Pearson Education Ltd 2019

General introduction

Paper 1 proved to be a reasonably standard paper in terms of difficulty, where the most common issue candidates faced was a lack of precision in their answers.

Question 1

Part (a) saw numerous mistakes made. Most commonly, candidates failed to define their population clearly - there were many instances of labelling gym members from 000-049 despite the fact that a **sample** of 50 gym members was required, not the starting **population** was 50 gym members. Some candidates had also rushed into answering the question and gave a description of how to get a **generic** stratified sample - in this instance it was specifically requested **half** the sample were 60 and over. As a general theme of the paper, reading a question carefully and fully should be emphasised.

A candidate who identified the correct issue in part (b) was likely to identify a solution to the issue in part (c). Many candidates were able to identify that more questionnaires should be sent out, but lacked the precision to gain the second mark.

Question 2

The first half of this question was done extremely well in general - most candidates were able to identify that the spreadsheet itself could be used if he had all data there, and calculation of correlation and regression on the calculator was done well. Interpretations should be done in context, but most candidates were able to correctly earn the marks here.

In part (d), however, many candidates suggested it could not be used as Liechtenstein is not in the sample rather than the fact that it was extrapolation. Part (e) could have been easily explained by pointing out that correlation and causation are not equivalent, though there were some excellent explanations of how population could be a third, confounding, factor.

Question 3

The calculations in part (a) were generally done correctly. In part (b) many candidates incorrectly converted 2 minutes 25 seconds to 2.25 minutes, scoring a maximum of 2 marks.

In part (c) the most common answer was that a low probability supported Rhodri's suspicion. However, in this particular context the event of 2 minutes 25 seconds passing occurs many, many times per day. It is important to think **carefully** about the specific context given.

Parts (d) and (e) had most candidates able to identify the correct causes of the variation. However some candidates did not pay attention to scale and considered that the variation in April was due to nights and scored 0 marks.

Part (g) had most candidates spot the lack of a constant average rate. Some candidates offer the scale as a reason as to why the Poisson would not be suitable and scored 0 marks. Some candidates correctly identified a lack of independence but if this was not justified with a reason, they would again score 0 marks.

Question 4

Candidates are reminded that bullet points should be used when presenting a long multi-part answer, each of which should be written in clear, specific, and concise sentences. Most candidates were able to score a good number of marks here, but five separate points were needed for full marks. Some candidates mentioned blocking, but again lacked the detail of exactly what blocking factors should be taken into account. Once again, precision is needed in written questions.

Question 5

Part (a) was a struggle for many candidates. **Queries** are part of the course, and candidates should expect more questions of this form in the future. In part (c) candidates had to be careful to treat it as the sum of 10 independent variables, rather than 10 times a specific variable - a maximum of 4 marks was available if this mistake was made.

In part (d) the precision surrounding the context could be lacking. Candidates might identify that all the vehicles were new, or the location was different, but then lack the detail when explaining **why** that was an issue.

Question 6

This question was a challenging use of Bayes' theorem. Some candidates had difficulty working out what tree diagram would be useful. However, this proved a good indicator of which candidates were well prepared.

Question 7

Part (a) was a standard application of approximating a binomial distribution with a normal distribution. There was nothing out of the ordinary in this question - a well prepared candidate should have been happy to see this question.

Part (b) was very challenging. Again, a level of precision was needed when answering this question - many candidates were able to identify you were less likely to correctly identify this proportion of cards than win the lottery - but missed out that this was on a **single selection**. Likewise, if discussing dice candidates should have mentioned that you were **only** making 31 throws.

Most candidates were able to answer part (c) correctly.

Question 8

In a **show that** question, a high degree of justification is required. Some candidates were lacking in detail for part (a)(i). In part (a)(ii) a diagram may have also been useful though candidates were not explicitly told this - the previous part of the question hinted that diagrams would be useful.

Part (b) required candidates to use the probability that was already given in part (a)(i). A candidate who had answered part (a)(ii) incorrectly was unlikely to score any marks for part (c). Part (d) was generally well answered, with candidates aware that trains don't always come on time.

Summary

Based on their performance on this paper, candidates should:

- be sure to use precise language and answer the exact question asked.
- read the question carefully and fully before answering the question.
- use **bullet points** with clear, specific, and concise language for explanation questions.
- try to get into the mindset of the context, rather than viewing the exam as a series of maths-led, routine questions.
- remember that not all explanations are statistical but may require some basic general knowledge and understanding in places.
- write conclusions to hypothesis tests in terms of **evidence**, rather than as a definite conclusion.

