



Pearson
Edexcel

Examiners' Report
Principal Examiner Feedback

Summer 2022

Pearson Edexcel GCE
In Statistics (9ST0)
Paper 01: Data & 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 2022

Publications Code 9ST0_01_2206_ER

All the material in this publication is copyright

© Pearson Education Ltd 2022

General Comments

This paper was accessible to most candidates, though most questions tended to also include some more challenging aspects that allowed more able candidates to demonstrate their knowledge. A common theme across the paper is use of context; in general it is not sufficient to simply state the assumptions required for a test to be valid, but rather how those apply to the situation in the question.

Report on Individual Questions

Question 1

In part (a) candidates were generally able to obtain the median and interquartile range or range from the cumulative frequency diagram and box plot, though some candidates incorrectly used the maximum frequency shown on the y-axis as the total number. Some candidates incorrectly referred to these as the mean and standard deviation, neither of which can be obtained from the diagram. When asked to compare spread a large number of candidates were providing comments comparing maximum and minimum values, rather than either measure of spread. Candidates should be reminded that comparisons should be done in context, referring to the average journey times from Stevenage and Wimbledon, rather than simply the averages from Figure 1 and Figure 2.

Part (b) was difficult, and in part (c) a large number of students had not read all the information, and gave comments about different train journeys. As all the commuters were on the same train from Wimbledon to Waterloo, there is no variability in that part of the journey.

Question 2

In part (a), many candidates only gave only one reason. The number of marks on a question is a good indicator of the depth that should be given when answering.

Candidates generally answered parts (b), (c) and (d) well, though many struggled with (e). It is important to note the difference between what is **impossible** and what is simply **unlikely** or unexpected; although a negative correlation in the population is extremely unlikely, especially when compared to the given data, it is not impossible and hence Alana is incorrect. It was encouraging to read responses from the most able candidates who indeed realised this.

Question 3

Candidates are advised to use bullet points when answering a question like this to be certain they have made enough points. There are some straightforward ideas that candidates should be familiar with when it comes to planning an experiment like using a large sample, which were not always mentioned. A large number of candidates went into great detail regarding how to generate a random sample, which was not the question that was asked. Many candidates did not address all three points, with some not mentioning what his sample should be doing, or suggested that Leo should ask people to rate how good they thought the paintings were rather than asking them which they thought was painted by a professional. There were, however, some excellent responses seen.

Question 4

This question had several parts that most candidates did well on – (a), (b), (e) and (f) were well answered. Parts (c), (g) and (h) demonstrated some of the harder aspects of the normal distribution – sample mean, a hidden binomial and conditional probability. Candidates would do well to be prepared for these, as standard probability calculations are straightforward due to the calculators available and therefore do not gain a large number of marks. In part (i) many candidates identified either the discrete nature of the data, or the upper bound as a reason the normal distribution was not appropriate, but did not provide sufficient comments for 3 marks.

Question 5

Part (a) was well answered, though a large number of candidates forgot their continuity correction so didn't gain full marks. Most candidates were able to justify the use of approximating each binomial to a normal distribution, but struggled to realise that the combination of the two normal distributions was part of Tom's approximation. In part(c), most candidates answered well, but should again be reminded to give answers in context; they were asked to justify why his assumptions may not be true so simply saying "may not be independent" is not enough as they have been told independence may not be likely, and are being asked **why** that is the case in this scenario.

Parts (d) and (e) involved some basic algebra. It is a good reminder of the additional skills in appendices 3 & 4 of the specification.

Question 6

This question was answered well, although some candidates did not provide contextual reasons as to why assumptions were not reasonable, simply saying that loss of one hair compared to another may not be independent. It is important to challenge the assumptions in context, rather than simply just restate the assumptions required for a binomial distribution, when candidates are asked to comment on the assumptions.

Question 7

Parts (a) and (b) were well answered, but part (c) had lots of possibilities for mistakes. Some candidates forgot to take into account that there was no replacement in this scenario, and some either calculated the wrong number of orderings, or forgot about them entirely. In part (d), (i) and (ii) were well answered, but many candidates gave the same answer for (iii) as they had for (i), missing the fact that the sampling with replacement was the important factor here. It is a good idea for candidates to make sure all the information provided has been considered. Parts (e) and (f) were better answered by candidates who chose to use a tree diagram than those who attempted them purely algebraically. It was pleasing to see some full mark responses on this question.

Summary

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

- give all explanations in context.
- look out for words shown in bold type.
- keep their working to more than 3 significant figures accuracy, throughout only rounding their final answer.
- use bullet points, each written in clear, specific, and concise sentences for explanation questions.

