Pearson
Edexcel

Examiners' Report

Principal Examiner Feedback

Summer 2025

Pearson Edexcel GCE
In Mathematics (9MA0)
Paper 31 Statistics

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 9MA0_31_2506_ER*

General Comments

The standard of work seen on this paper this year was encouraging with most learners being able to make a start on all questions but some of the later parts such as 2(c), 4(e) and (f) and 6(c) provided some discrimination.

Comments on individual questions

Question 1

Most learners appeared comfortable working with tree diagrams and could find the probability of selecting a yellow bead from bag **B** correctly. A small minority wrote down a correct equation such as $\frac{4}{12} \times p = \frac{3}{16}$ but did not solve it correctly but the vast majority obtained $p = \frac{9}{16}$ and were able to go on and complete the tree diagram correctly. A few responses had different pairs of probabilities on the branches for bag **B**.

Part (b) was usually answered well, but the most common error was to overlook the "at least" condition and omit the case for two yellows. The most popular approach was to identify the 4 cases from the tree diagram but a few used $1 - \mathrm{P}\left(\overline{Y}\,\overline{Y}\right) = 1 - \frac{2}{3} \times \frac{7}{16}$ and some used $1 - \mathrm{P}(RR) - \mathrm{P}(GR)$. Few responses did not present their answers in (b) and (c) as exact fractions as required.

In part (c) many could write down $\frac{\mathrm{P}(W \cap X)}{\mathrm{P}(X)}$ but appreciating that $W \cap X$ meant that the case GY was required defeated a large number. It was fairly common to see the answer to part (b) or $\frac{9}{16}$ appearing in both the numerator and denominator and leading to an incorrect answer of $\frac{1}{4}$.

Question 2

The majority of learners answered parts (a) and (b) very well. A very small minority used $\frac{364902}{6612}$ in part (a) and a sizeable number found the variance not the standard deviation in part (b).

Part (c) proved to be more challenging with many choosing coach A on the basis that the lower standard deviation meant that those runners would be more "consistent". A good number of responses chose coach B citing the larger standard deviation as a reason, but few were able to explain that the larger standard deviation increases the likelihood of an extreme value such as the fastest time. A small minority of responses did not consider standard deviation and instead incorrectly focused on the different sample sizes.

Question 3

Part (a) was usually answered well. Some explicitly stated the correct model but others achieved this mark by giving one of the correct probabilities though accuracy marks were often not gained as a number of responses truncated in (i) to give 0.0847 and others simply gave 0.002 in (ii) rather than to the 3 significant figures expected. In part (ii) the most commonly seen incorrect answer was 0.0167 caused by thinking that the phrase "more than 3" included 3 or using $1 - \mathrm{P}(D \leqslant 2)$ as the complement for $\mathrm{P}(D > 3)$.

In part (b) most responses had the correct binomial model or inferred it by stating a suitable probability. A common mistake though was to use $\mathrm{P}(X = 8)$ rather than $P(X \geqslant 8)$ which did not gain either of the remaining marks. Those with the correct probability sometimes gave a comparison with 0.05 rather than 0.025 and did not gain the final mark and others did not give a conclusion in context. Although a critical region was not required here (we will only ask for this if it can be found using the tables provided) some did use this approach. Many were successful but a few did not give both tails of the critical region and others did not state the critical region properly for example by just stating $\mathrm{P}(X \geqslant 9) = 0.0177...$ and not writing the critical region as $X \geqslant 9$ and $X = 0$. A small number attempted to use a normal approximation here.

Question 4

On the whole this question proved quite challenging as it tested a number of topics: sampling, regression, correlation and hypothesis testing as well as using the large data set and having a number of parts requiring a written explanation. However, there was something for all learners and very few blank responses were seen and parts (e) and (f) provided an opportunity for differentiation.

Part (a) was answered well though some incorrect responses of "stratified" or "quota sampling" were seen. There were very few responses not using the correct terminology here. In part (b) most identified that "trace" or "tr" would need attention, but some did not link this to the rainfall variable. In part (c) (i) there were many correct answers and some who identified mm or hours but not in a correct combination: answers such as mm^2 or even $\text{hour}\,\text{mm}^{-1}$ were seen, whilst others appeared not to be that familiar with the large data set and used other incorrect units such as oktas, lumens or Hpa. Most responses gave a suitable contextual response to part (ii) but some simply said that 0.714 was the minimum rainfall rather than the (average) rainfall when there was no sun. The majority of responses recognised the inconsistency between Kay's teacher's claim and Kay's data and many cited "positive correlation" for Kay's data as their reason. A few opted to substitute two different values into the equation of the line to demonstrate their point. A small minority believed the teacher's claim to be true and they were perhaps relying on their knowledge of weather conditions rather than interpreting the data Kay had collected.

Part (e) was found to be much more challenging for learners. Many were able to state the hypotheses correctly although some had an alternative hypothesis of $\rho > 0$ and a few had a null hypothesis of $\rho = 0.015$ though few were using r instead of ρ. Very few responses interpreted the p-value with many comparing it to 0.16 or comparing 0.16 with 0.05 but some gave a correct conclusion in context. Good attempts to part (f) were rare and blank responses not uncommon. Many answers did not give full and precise definitions and others lacked clarity so that they could have been interpreted as either the sample or a population. For part (i) the information was given in the question but few were able to extract the relevant information: the variables (sun and rain) for May to October (or from the large data set) from 2015. Part (ii) produced a slightly better response with successful comments such as "sun and rain from all of 2015" or "all the variables in the large data set" or "from all across the UK" being seen. Common mistakes often involved describing a population for which the sample was not a subset such as using months from April to November or the year 2016.

Question 5

This question was answered very well. Most knew how to use their calculators to answer part (a) though some didn't appreciate the difference between discrete and continuous random variables and in part (i) calculated $\mathrm{P}(M \geqslant 187)$ whilst others incorrectly applied a continuity correction.

In part (b) many correctly formed and solved their simultaneous equations, but some made sign errors in their z-values and a few did not standardise correctly for example using 0.4 and 0.15 as z-values. Those who sketched a diagram of the distribution usually avoided sign errors. Some did not use accurate enough z-values (the tables give values to 4 decimal places) and this often meant they lost an accuracy mark for their value of sigma. Many relied on their calculator for solving the simultaneous equations which was fine if the equations were correct but without some written attempt to eliminate one variable, we were unable to give further credit if the equations were incorrect. Some responses accepted a negative standard deviation without checking to identify the error.

Part (c) was completed successfully by many learners though many used their calculated mean and standard deviation from part (b) to calculate $\mathrm{P}(170 < F < 175)$ rather than using the information in the question in the calculation $1 - (0.4 + 0.15)$. A few responses added the two probabilities rather than multiplying here.

Question 6

This question provided some challenge and discrimination for the end of the paper. Most learners attempted this and made some progress.

Part (a) was a "show that" question and learners need to remember therefore to show all steps in their working clearly and present a solution without incorrect working. Most were able to identify that they needed to use the fact that the sum of the probabilities equalled 1. A popular starting point was to write down $1=\frac{2}{k}+\frac{4}{k}+\frac{6}{k}+...+\frac{2n}{k}$ and some were then able to recognise that an arithmetic series was involved and use a suitable formula to find the sum. Some learners didn't use the correct number of terms with $2n$ or 4 (from the 4 terms 2, 4, 6 and $2n$) being common errors. Full and concisely presented solutions were rare.

Part (b) was generally answered well, though some were confused by the even number restriction and attempted $\frac{16+17+18+...+24+25}{420}$ and a few included 26. Those who attempted $\mathrm{P}(R \leqslant 24) - \mathrm{P}(R \leqslant 14)$ were rarely successful but many simply listed the 5 cases required and were readily able to reach the correct answer.

The attempts seen in part (c) were encouraging with many realising that the discriminant of the quadratic equation was required and for "no real roots" this must be negative. Most of these learners were able to form the correct quadratic equation in g and find the correct critical values but the correct critical region eluded some with $g < 2$ and $g < 10$ being a common error. From the correct inequalities for g it was a small step to identify that the values of g required were 4, 6 and 8 and the majority of learners who got this far were able to complete the solution and state the correct probability in the correct form.

Pearson Education Limited. Registered company number 872828
with its registered office at 80 Strand, London, WC2R 0RL, United Kingdom