



Pearson
Edexcel

Examiners' Report
Principal Examiner Feedback

Summer 2022

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 2022

Publications Code 9MA0_31_2206_ER*

All the material in this publication is copyright

© Pearson Education Ltd 2022

It is pleasing to see that students were attempting all questions with few completely blank papers. They are becoming adept at using their calculators to find probabilities and standard statistics but must be aware that for a "show that" question they will need to show their methods. There is evidence of progress in their familiarisation with the large data set where a good knowledge of the variables and how they are recorded and familiarity with where the places are in the world gave some easy marks. There are still areas of weakness in laying work out clearly. This was particularly evident in Question 2 where the work became quite muddled with no labelling or reference as to what they were trying to work out and often led to errors.

Question 1

This was generally a good start to the paper for many students with almost all being able to score some of the marks on this question.

In part (a) students were usually able to find at least one of the two required probabilities although occasionally they miscopied from their calculator with 0.197 being written down rather than 0.0197. The notation demonstrating use of a binomial distribution was varied. Some students wrote down answers without stating the distribution that was being used, in this instance this was condoned provided that a correct probability was seen, but students should be aware that method marks are often awarded for sight of the correct model. Where full marks were not awarded the most common reasons were, rounding answers to fewer than 3 significant figures, not knowing whether to use Binomial PD or CD on their calculator, or for not knowing how to rewrite $P(X \geq 5)$ in a form that allowed use of the binomial CD function.

Part (b) caused more problems for students. The question clearly asked the students to use a normal approximation and most found the correct mean and variance for the Normal distribution and went on to find a probability. There is some confusion about whether it is the variance or the standard deviation which is a parameter of the normal distribution and too many failed to summarise their working with a simple statement $N(120, 64.2)$. Almost all students correctly interpreted "more than 110" as requiring $P(X > 110)$ but a minority of students remembered to use a continuity correction, required when approximating a discrete distribution with the normal distribution, and those that did attempt to use a continuity correction often used an incorrect one, most commonly 109.5. Despite being directed to use a normal approximation by the question some students used the binomial distribution. This gave an answer very close to the one obtained from the correct distribution, so it was very important to write down the distribution used, the probability statement with its continuity correction and give the answer to a sufficient degree of accuracy before rounding.

Question 2

Students had mixed success with part (a) of this question. Most students who recognised that they needed to use standardisation and found a z-value of ± 1.96 completed the proof without error scoring both marks. A few students who used + 1.96 then "fudged" the signs part way through their solution losing the A mark.

Part (b) was a good introduction to part (c) with the majority of students scoring the mark although there were some who didn't read the question carefully enough and used the 7.902 from (a) as the lower limit rather than 7.94

The majority of marks lost were due to rounding to 2 s.f

Most students were able to make a credible attempt at answering part (c). However, there was a lot of information for students to process and it was common to see errors or omissions at some point in their solution. There were several correct ways of approaching this question. All approaches required the calculation of the probability of a rod lying in each of the categories, some students mistakenly used $P(L < 7.94) = 0.025$ presumably assuming that 7.94 was the value given in the question (7.902). Another common error was assuming that 7.94 and 8.09 were evenly spaced about the mean so the probability from (b) was subtracted from 1 and then the result halved.

Once the students had found the probabilities a common approach was to start with calculating the number of rods in each category, and then calculating the income/profit. Many students calculated the total income and went no further. Whether this was because of a misunderstanding of the term “profit” or whether having got that far they simply forgot to subtract the costs was unclear. The ability for students to apply their skills in “real life” situations needs further development including training in setting out working methodically

Common errors included rounding the probabilities inappropriately, rounding the number of rods to integer values and some occasionally being uncertain whether they were working in pounds or pence. The most successful answers were those where the calculation was combined into a single expression.

Students who attempted part (d) usually selected the correct Binomial model. A few students, recognising that n was large, used a normal approximation which was not appropriate in this case with very small p . It was common for students to find $P(X \leq 6)$ rather than $P(X < 6)$. Those who made this error were however still able to access the mark available for a correct interpretation. A minority of the students who gave a correct interpretation did not make the required comparison of their probability with the given 95% aim thus losing the final A mark. Students need to be aware that in questions such as this they should be using the values calculated to justify and explain any conclusions they make.

A reasonable proportion of students were confused about what the question was asking and tried to set up a hypothesis test, making a conclusion based on whether they were accepting or rejecting the null hypothesis rather than judging the likelihood of the manufacturer meeting his aim by comparing their probability with the given 95% aim. Some Weaker students simply worked out 6 out of 200 as a percentage and compared this with 5%.

Question 3

Part (a) tested knowledge of the data set. Some students were able to identify or trace as the required answer, but many displayed their lack of knowledge of the data set, with n/a being a typical wrong answer, or 0 or 0.025. Many students gave no answer of any sort and there were a few nonsensical or confused answers such as explaining how the rainfall might be collected (a particular favourite being ‘with a bucket and ruler’) or how it might be recorded (in an excel spreadsheet).

The vast majority of students were able to find the mean correctly in Part (b). Mistakes if any in (i) tended to be rounding the answer to 5.6 without previously showing a more

accurate answer, hence losing the single mark. A few students showed the calculation without evaluating it. Disturbingly there was a tiny minority who seemed to have formulae involving square root signs etc. It is disappointing to see evidence of students studying statistics at advanced level who still do not know how to calculate a mean. There were more errors (ii) although well answered by the majority. Errors tended to be missing the square root in their expression for the standard deviation giving the variance as the final answer, with students believing this to be the standard deviation; subtracting the mean rather than the square of the mean premature approximation of the mean leading to inaccurate final answers and giving the answer to 2sf or an even lesser degree of accuracy. A small number of students gave an answer with no working and were then at the mercy of having obtained a sufficiently accurate answer or risk losing these relatively easy marks. The calculation and use of S_{xx} was seen occasionally.

Part (c) proved to be more problematic. Many students lacked knowledge of the LDS and did not know the relative geographical locations of Camborne and Leuchars within the UK. As such, a significant number of students made the incorrect assumption that Leuchars is in the South and Camborne is in the North. However, when students were able to locate the relative positions of the two places correctly, they were generally able to draw the correct conclusion, resulting from comparing the mean rainfall between the two places.

The main errors included

- not making reference to the relative positions of the two locations in their answer e.g. stating that Camborne was in the South with no reference to the relative position of Leuchars,
- not making any statements about either of the geographical locations at all and only compared means, with responses such as: no, as the mean is less in Leuchars than in Camborne.
- not drawing the correct conclusion after correctly comparing locations and means
- not stating any conclusion.
- incorrectly comparing standard deviations rather than means.

There was a wide variety of responses to part (d). Successful students often stated one or more of the conditions of the Binomial Distribution, with reasons why this distribution was not suitable in this case. e.g.

- the probability needs to be independent but the probability of rainfall is not independent as it often depends on the weather from the day before.
- the probability of rain needs to be constant but the probability was taken from May to October and summer is June to September and therefore the probability is likely to be higher than in the summer when there is less rain

Errors included

- referring to independence without context
- commenting on the great storm of 1987 or global warming
- stating that the data is out of date
- the sample size is too small
- some values are trace values and/or rainfall is influenced by other factors.

Question 4

Part (a) was well answered with the majority of students writing their hypotheses in terms of p .

Part (b) proved to be a challenge for many of the students with a reasonable proportion unable to even begin, which is unusual for a question of this nature – as it is quite standard.

Most students were able to state the correct distribution and write some probabilities, but many were not able to identify correctly the CR from their probabilities.

Responses scoring M1A1A1 were rare. This was often due students giving probability statements rather than giving a CR

Other issues included:

- giving probabilities to only 2 s.f.
- selecting the region between the two Critical values as their CR

Students who were aware of what a two-tailed CR looks like almost always got part (c) correct. The students who did not really understand the idea of the correct CR seemed to make a fairly random guess at the actual significance level.

For those students who stated both tails of the critical region in (b), the majority correctly compared 15 with their critical region and then usually supported the manager. Others used the longer version of the contextual statement and most did so correctly.

Question 5

Part (a) was well answered with most students able to gain the mark for a correct answer. A simplified fraction or the correct decimal value was often seen.

Whilst the majority of students continued to score well in part (b) the most common incorrect answer came from students who thought the question was conditional, giving $\frac{170}{550}$, i.e. the probability of not being a professional *given* they live in area B . Some students attempted a product of probabilities assuming independence, which was not given and not correct.

In part (c) there was mixed success in the interpretation of the information given in the question with only about half of students getting the Venn diagram completely correct. Some correct diagrams followed minimal working whilst others filled at least a page. A common omission was to leave the region $H' \cap R' \cap F'$ blank. The method mark was most commonly given when implied by one correct value in the Venn diagram, rather than for the student's working, which was commonly unclear. Some working indicated incorrect use of the given percentages. Where only one value was correct in the Venn diagram it was usually the 481 or 259, for 'professionals from area A '. When none or only one of the four correct values was seen many students benefited from being able to achieve 1 mark for the sums of certain pairs of values in their diagram. There were a number of students who simply abandoned the Venn diagram which often resulted in no attempts being made at the remaining parts of the question.

Part (d) had a mixed response with many correct answers. This part could be answered independently of whether or not (c) had been attempted and it was good to see most students doing so successfully

Whilst many realised the 133 given in the Venn diagram was required, the follow-through mark could only be scored by students who had some value in the region $H' \cap R' \cap F'$ so that this value could be added to the 133 in their calculation. A common non-scoring answer of $\frac{133}{1825}$ was often seen.

Most students realised that part (d) was a conditional probability question with many taking the simpler approach of using numbers rather than probabilities in their ratio. Those with a correct Venn diagram usually scored both marks here. As is often the case, a noticeable number of students had the incorrect event, namely F , as the denominator in their probability expression rather than H .

Question 6

In part (a) a large number of students simply wrote 'negative correlation', with no context. The word Interpret in the question hints that context is required. Of the students that did try to truly interpret the correlation in a contextual statement, the vast majority were successful.

Whilst many students realised that one-tailed hypotheses were needed in part (b) few used ρ for the parameter. Most students had \pm the correct CV but the final A1 was lost by many students, as they compared the positive value 0.3887 to the negative test statistic. This showed these students were unfamiliar with the meaning of the CV/CR and how it related to their test statistic.

Part (d) differentiated well between students with basic knowledge and those with deeper understanding.

The majority of students scored 5/5 on this part. Almost all used the LH approach on the MS. This was usually well done, though a good number made errors with the laws of indices and maintained a + between the two terms after making the equation into powers of 10. The students knew their log rules, but many did only so to a basic degree, and came unstuck during their working. Sometimes this was to such an extent that they lost the marks.

A surprisingly high number of students used natural logarithms and hence the exponential function in this question showing a less secure grip of this topic.

