Pearson Edexcel

Examiners' Report
Principal Examiner Feedback

October 2020

Pearson Edexcel Advanced Subsidiary
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

October 2020
Publications Code 9MA0_31_2010_ER

Despite the much reduced entry for this paper there was evidence that it was more accessible than June 2019. The questions were graduated well with the most challenging questions being 4 and 5. There were however a number of weak candidates who did not appear to have prepared thoroughly for this examination.

Comments on individual questions

Question 1

This proved to be a popular start to the paper with over 35% of students scoring full marks. Part (a) was answered well with most understanding the term "mutually exclusive" and giving their answer as events not probabilities. Part (b) was very straightforward and nearly all could interpret the Venn diagram correctly to answer this. Many knew a suitable rule for independence in part (c) and could use this to find P(A) = 0.6 or write down a suitable equation for q which was then solved correctly. Part (d) was more challenging as many could not use the given conditional probability correctly to obtain an equation for r but most knew how to use this answer to find s and were able to gain at least one mark in this part.

Question 2

Very few students were unable to identify the negative correlation in part (a) with many attempting to describe the strength as well (split fairly evenly between weak and strong) though this was not required here. In part (b) it was clear that more students were familiar with the large data set. Many identified "rainfall" as the variable on the y-axis and they usually gave the correct units although some thought it should be mm^3. The most common incorrect answer here was "cloud cover" but, of course, this variable is not available on the large data set for Perth. Part (c) was a standard hypothesis test for correlation. It was encouraging to see a good number using ρ for the hypotheses but fewer realised that the test was two-tailed. The phrase "…believes there is a correlation between…" shows that a two tailed test should be used, however many were swayed by the negative value of r and opted for a one tailed test. It was good to see that most chose the critical value from the tables that was compatible with their alternative hypothesis and this meant that they could also access the final mark provided their comment was in context, "support for Stav's belief" was perhaps the simplest way of including the context. There were some good answers to part (d) but many others were unable to pull together the various parts of the question to make a correct inference. Some stated that the humidity was high and other assumed this but interpreted the negative value of r correctly and were then able to state that they would expect the number of hours of sunshine to be low. Some gave a suggested value for the number of hours of sunshine but this needed to be realistic and values more than 5 were not accepted. A few students knew (from the large data set) that humidity greater than 95% implies foggy conditions and argued correctly that this would suggest a low amount of sunshine.

Question 3

Part (a) should have been an easy start to the question but a surprisingly large number of students gave an answer of 33 by failing to consider the outliers to determine the range. Part (b) though was answered very well and nearly all students answered (c) correctly. The presence of the formula for standard deviation in the formula booklet meant that a good number answered part (d) correctly although occasionally some forgot to take the square root of failed to square the mean. In (e) most gave a suitable calculation to support their statement that Taruni would identify only 1 outlier. Some candidates used the median of 20 rather than the mean of 22.5 though. The last two parts were more challenging but it was encouraging to see a good number of attempts. Many were able to deduce that because the mean did not change then $a + b = 45$ but fewer were able to establish that, because the median increases, both values had to be greater than 20. Some did state these two conditions but sadly did not choose suitable values ($a = 25$ and $b = 20$ was a common error) for their final answer. In (g) many students stated that the new values would be "close to the mean" but this was not precise enough to ensure that the standard deviation would decrease. Some of the very strongest candidates did realise that because both values were within 1 standard deviation of the mean then they could be sure that the new standard deviation would be smaller. This appreciation that standard deviation represents "average distance from the mean" is clearly not well known.

Question 4

Part (a) was generally answered very well. Being a "show that" question meant that we needed to see a full explanation and most clearly showed the sum of the probabilities equal to 1 and then solved the resulting equation to find *k*. Failing to show this key step usually meant that no marks could be awarded. A significant number of students left the rest of the question blank but those who did make an attempt were usually able to identify some, though rarely all, the cases in (b). Finding the probabilities was more of a problem with some using $P(D = 50) + P(D = 30)$ rather than the product and those having correct probabilities for all 3 cases were quite rare. It was good to see many students who had faltered on (b) still making a start on part (c). Many identified the 4 angles as terms of an arithmetic series and some were able to reach the condition $2a + 3d = 180$ (or equivalent) but many could not see how to complete the problem. Some used the information that $a > 50$ to reach $d < \frac{80}{3}$ but were unable to make the connection that this meant that only the two cases $d = 10$ and $d = 20$ were required and so the probability could simply be written down.

Question 5

It was disappointing that over 20% of candidates scored no marks here as there were some straightforward marks available to those who had studied this part of the specification. Part (a) was a very straightforward opening part and the vast majority of students could use their calculator efficiently to score this mark. Part (b) was a fairly standard hypothesis test and a good number made progress here. Some still did not use μ for the hypotheses and a null hypothesis of $\bar{T} = 10$ was all too common. Most attempts used a normal distribution with a correct variance but a significant number of students used a mean of 11.5 rather than 10. This is not a correct method for a hypothesis test (the probabilities should be calculated on the assumption that H_0 is true) but, on this occasion, for those who still managed to reach a correct conclusion a special case enabled them to secure one mark. A variety of different approaches were used in part (b), some finding the critical value for $\bar{X}$ and others calculating the test statistic of 1.677… but some of these candidates compared this value to the significance level rather than the critical value of 1.6669 from the tables. The first part of (c) was another fairly straightforward calculation which many scoring this mark, sometimes including those who failed to score the mark in part (a). Students need to be aware of which "tail" their calculator gives them. Beyond (c) (i) the question was more challenging but it is encouraging that over 50% of students scored some of these first 6 marks.

In (c)(ii) although a good number of students recognised that a conditional probability was required, few could find a correct expression for the numerator and a common error was to see $\frac{\text{(c)(i)}}{\text{P}(T > 0)}$ used rather than a correct numerator of P(0 < T < 2). In part (iii) many realised that the model would assign some probability to negative values of T but this will always be true when a normal distribution is used as a model. The point here was that a significant proportion of the values for T were negative and so this means that the normal model is not very good. Most responses seemed to believe that a model must match reality exactly but a few did comment that P(T > 0) should be closer to 1 or that the difference between (i) and (ii) was too big and therefore the model was not a good one to use.

Fully correct answers to part (d) were rare and this part gave the A* students an opportunity to shine. Successful students realised that they needed to look at P(T > 2) and often found this by taking 1 – (c)(i) but they rarely realised that the median splits this probability into 2 so they were seeking the value of t such that P(T > t) = 0.40…Some who reached this stage tried, incorrectly to use their calculator, failing to realise that their calculator gave t for probabilities of the form P(T < t), a few used the tables provided to find z = 0.2533 and some did prepare correctly to use their calculator and found P(T < t) = 0.5978… (the 0.402… + (c)(i)) from whence the correct value of t = 5.9 could be found. It was encouraging to see some good use of diagrams in this question being used to help students identify which probabilities they needed.

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