Examiners' Report

Principal Examiner Feedback

Summer 2022

Pearson Edexcel GCE

AL Further Mathematics (9FM0)

Paper 4B Further Statistics 2

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 9FM0_4B_2206_ER*
All the material in this publication is copyright
© Pearson Education Ltd 2022

Introduction

Candidates were fully prepared for most of the topics on this specification and appeared to make good use of the Advance Information provided for this series. Modelling assumptions are more well-known than previously but candidates still need to work on engaging with the context of the questions (rather than relying on repeating ‘textbook’ answers).

Question 1

This question proved to be an accessible start for virtually all candidates. Nearly all made a good start to the paper in part (a) by being able to calculate the output of the linear regression equation. Part (b) was generally well attempted with the majority of candidates giving a correct description of a residual. Incomplete answers simply stated how far away from the line of best fit without referencing the vertical/*y* distance.

Part (c) was again well attempted, though many lengthy attempts were seen at calculating the product moment correlation coefficient when candidates should have realised that they had all of the information to complete it in one step. It was not uncommon to see 2 significant figure answers given rather than 3 as instructed on the front cover of the examination paper.

(d)(i) was well accessed with most stating that there was weak correlation. Stating that the value of r was far from one was insufficient as candidates needed to make it clear that it is the proximity to 0 that makes it weak. Candidates struggled more in part (ii) to clearly express themselves, though a number of candidates did correctly identify that for values of $t > 20$ the residuals were no longer randomly scattered about 0.

Part (e) was the least successfully attempted part of the question with the majority of candidates suggesting that Kwame was indeed correct. Most did not realise that the value of RSS is not ‘standardised’ and depends on the size of the variable being measured.

Question 2

Part (a) was extremely well answered with most selecting the appropriate z-value and going on to find the correct confidence interval.

Part (b) was less familiar to candidates and correct answers were rarely seen. In part (i) many candidates gave the distribution of $\bar{Y} - \bar{W}$ using the information from the sample data rather than what its distribution under the null hypothesis; that is to say that most candidates thought that the mean was 0.3 instead of 0. In part (ii) many thought that the CLT allows us to assume that the distributions of Y and W are normally distributed when in fact we don’t need to know anything about their distributions.

In part (c), many fully correct responses were seen with correct notation in the hypotheses and an accurate critical value stated from the statistical tables. Only on rare occasions were less accurate critical values used and conclusions not given in context.

Question 3

Candidates are generally well prepared for this type of question and most successfully scored full marks here. Occasional slips included forgetting to square *b* in the equation for the variance. Most were able to solve their simultaneous equations with ease and go on to select the appropriate pair of solutions from the quadratic equation.

Question 4

The candidates who recognised that a paired *t*-test was needed knew what to do. The errors made were usually minor such as making a slip with one of the differences (usually the last one) or failing to match their alternative hypothesis with the sign of their differences. Pleasingly most candidates selected the appropriate critical value for their *t*-value. The most common error was to use a difference of means test which resulted in those candidates gaining a maximum of 1 mark for identifying an appropriate critical *t* -value. In part (b) only a few candidates showed understanding that the differences in the resting heart rates need to be normally distributed.

Question 5

In general, candidates did not engage with the context of the question in part (a) and many lost the mark for answers which were too vague. Candidates should be made aware that just saying “it must be normally distributed” will not gain the mark.

Part (b) was generally well attempted. One common error was to only find one tail of the critical region. Another common error was to write the word ‘and’ in between their two sets essentially stating it was the intersection of the two regions.

Part (c) had a greater success rate and many fully correct responses were seen.

Question 6

This was one of the most demanding questions on the paper and fully correct responses were only given by the most able candidates. Part (a) was a ‘show that’ meaning that candidates were supposed to show all stages of their working and give the conclusion that both of the estimators are unbiased. Many skipped steps in working out the expectation of *K* and did not make it clear they were using the sum of an arithmetic series. Poor notation was common as many did not use expectation notation until the end of their solution.

(b)(i) was challenging for many. Many recognised that the sum of the variances was needed, but most did not recognise the need to use the sum of the squares. On some occasions the square of the sum from part (a) was used. Some correctly factorised out $\frac{2^2}{n^2(n+1)^2}$ but made no further progress.

(b)(ii) had a higher success rate with most showing some understanding of both $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$ and $\mathrm{Var}(X_1 + X_2) = 2\mathrm{Var}(X)$ but some could not correctly simplify their answer to a single term.

In part (c) many showed knowledge that the unbiased estimator with the smallest variance was the better estimator, but only a few were able to fully justify for large values of *n* the limit of each of the estimators. This part again discriminated the most able candidates on this paper.

Question 7

Another challenging question at this stage of the paper, but it was pleasing to see many correct responses. The question gave a big clue stating that integration must be shown – yet some candidates did not attempt any integration at all in their answers. Most recognised that $\mathrm{f}(l) = \frac{1}{6}$ and many went on to give a correct expression for the perimeter. A common error was to attempt to integrate the expression $l(2l + \frac{80}{l})$ rather than $\frac{1}{6}(2l + \frac{80}{l})$. Of those with a correct expression, the vast majority showed the correct integration with the substitution of limits in order to arrive at the correct answer.

Question 8

On the whole, this was the most demanding question on the paper, but parts (a) and (b) and to some extent (d) provided accessibility.

In part (a) most candidates equated the cumulative distribution function to 0.5 and went on to solve and select the correct exact value for the median. Some were penalised for giving an inexact answer whilst others believed the F(0.5) was required to find the median.

Part (b) was more difficult but most realised that a conditional probability was required with P(*X* > 1.2) on the denominator. The numerator proved more challenging and it was not uncommon to see P(*X* < 1.6) used.

Part (c) was the most challenging on the entire paper as virtually all candidates were unable to deal with the inequality in the probability $\mathrm{P}\left(\frac{1}{X} \leqslant y\right)$. Many simply replaced *x* with $\frac{1}{y}$ but did not consider the effect that would have on the limits. Many unsuccessful attempts at differentiation and integration were seen.

In the final part of the question (d) candidates should have realised the need to differentiate twice to see if there were any maximum turning points in the range. Some were able to make progress here despite the difficulty in (c) and marks were still available for those who persevered.