Principal Examiner Feedback

Summer 2018

Pearson Edexcel GCE Mathematics

In Statistics S1 (6683/01)

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 2018
Publications Code 6683_01_1806_ER

Introduction

Although this paper proved to be a little more difficult than some others in this series there were marks available for all students in each question. Questions 3, 6(c) and 7(b) and (c) proved particularly discriminating.

Comments on individual questions

Question 1

This proved to be a relatively friendly starter with nearly 62% of students scoring full marks. Most were able to make some progress in part (a) though a few confused the cumulative distribution function with a probability function.

Question 2

The first 5 parts of this question were testing ideas familiar to most students and typical responses were seen. In part (a) most could find the width correctly and many knew about frequency density but this was not always applied accurately. In part (b) the usual problems of identifying the class width or the class boundary were the main source of errors but many could find the median correctly. Most found the correct mean in part (c) but there were still plenty of students who did not know a correct formula for variance or standard deviation. In part (e) most were able to substitute their values correctly into the formula for skewness and most had a positive value here which they could use in part (f). Some of the best solutions to part (f) involved simple sketches of positive and negative skew followed by a comment about there being more longer delays on Friday. Some ignored the calculation they had just performed and the extra information given in the question about skewness on Friday and simply wrote about reasons for delays on motorways rather than commenting on the lengths of delays as intended. A few confused skewness with correlation.

Question 3

Many students found a correct answer to part (a) though some took a very long route to arrive at this. Part (b) proved more challenging with many students trying to standardise with $Y = 17$ but failing to attempt this with $Y = \mu - \sigma$ which was the key to unlocking this part of the question. There were lots of attempts to find an answer using the 0.2533 value from tables but this was often used as a probability not a value of z. Few students drew a diagram of a simple normal curve with significant values of Y and associated probabilities marked on it.

Question 4

This question discriminated quite well and there was a good spread of marks. In part (a) many started with a tree diagram giving probabilities in terms of y and r. To complete the problem though they needed to use the fact that $(r + y) = 63$ and many didn't seem to appreciate that it was possible to arrive at a numerical value for the answer to part (a). Of course some realised that picking green or (not green) followed by green would give the answer straight away. Most were able to prove the required result in part (b) and this was often solved correctly in part (c). It is important to remember that in "show that" questions such as this we need a full explanation and those who simply showed that $r = 16$ satisfied the equation had not shown that $r = 16$ was the only possible value. We did expect to see the – 15 solution mentioned and dismissed. In part (d) some didn't realise that a conditional probability was required but many did and gave a ratio of probabilities with the correct numerator. The calculation of the probability for the denominator was not well done: many tried listing the 5 cases *RG*, *GR*, *RY*, *YR* and *RR* and their probabilities but often at least one of these was missing. It is perhaps surprising at this level that more students did not attempt 1 – P(no reds) which gives the result very simply.

Question 5

Most students could make some progress with this question though part (a) was not always answered successfully. Some identified the symmetric pattern in the probabilities, whilst others mentioned that the x values were symmetrically placed around 2 but only a few identified the symmetric nature of the distribution as a whole. A popular alternative and successful route in part (a) was to form an expression in terms of a and b for $E(X)$ and then use the result about the sum of the probabilities = 1 to complete the proof. A number simply found the mean of the 5 values of x and a small minority stated that the distribution was uniform and they, obviously, failed to secure the mark. Whatever their score in part (a) most wrote down a correct equation in part (b). In part (c) many made a correct start and found $E(X^2)$, often correctly. Some thought that $E(X^2) = \text{Var}(X)$ but most went on to use a correct formula to reach a second equation in a and b. Solving their 2 linear equations in part (d) was usually started well but a variety of arithmetic errors in parts (c) or (d) meant that the correct probabilities were not always seen. In part (e) those who used the $E(aX + b)$ and the $\text{Var}(aX + b)$ formulae were usually successful, some chose to find the distribution for Y and sometimes made errors. Part (f) discriminated quite well. Those who solved the inequality or compared values for X and Y were often able to identify the correct values and with a follow through mark at the end could often score full marks here. Some found separate probabilities for X and Y and multiplied them together and made little progress.

Question 6

Nearly everyone was able to answer part (a) correctly and this provided a familiar start for all students. Part (b) was a fairly standard question too and many answered this correctly too though a small minority thought that a negative correlation meant that a linear regression model was not supported. In part (c) despite the form of the equation of the regression line being given many proceeded to find S_{th} / S_{tt} instead of the correct S_{th} / S_{hh} . Some realised that it would be easier to find the regression line of h on t and thought that they could then rearrange this into the required form which, of course, would not give them the correct answer. Those who did realise that they needed S_{hh}, and that this could be found from the given value of r, were usually able to go on and find a correct value for b and a though there were a number of arithmetic errors, and sign errors when finding a, that prevented some from obtaining full marks. In part (d) many failed to engage fully with the context and simply said that a represented the temperature when $h = 0$ rather than referring to the temperature at sea level and some thought it gave the height at which temperature was zero. In part (e) many simply found the temperature at a height of 150 metres rather than the drop in temperature. Those with an incorrect gradient in (c) seemed unperturbed by a temperature drop of 11520 here.

Question 7

Part (a) was a standard question on the normal distribution and most students were able to answer it successfully. Although we have had questions in recent papers involving conditional probabilities and the normal distribution parts (b) and (c) proved too challenging for many of the students here. In (b) some realised that the proportion Adam could sell was (1 – their answer to (a)) but they sometimes multiplied by 0.75 instead of 0.25. To make progress in part (c) they needed to identify a correct conditional probability and there were basically two options. Some opted for $P(W < q_1 \mid W > 92) = 0.25$ which gave a complicated numerator for their ratio of probabilities. An easier route was to use $P(W > q_1 \mid W > 92) = 0.75$ and this more often lead to a correct answer as the numerator is the more straightforward $P(W > q_1)$. Those who did calculate their conditional probability correctly often went on to find the correct answer but some did not consider the sign of their z value carefully enough and ended up finding the upper quartile rather than the lower quartile as required. In part (d) some realised that a probability of the form 0.25x0.25x0.5 was required but they rarely accounted for all the cases: multiplying by 3 rather than 6 was quite a common error.

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