Pearson

Examiner's Report
Principal Examiner Feedback

Summer 2018

Pearson Edexcel GCE
In Statistics (8ST0) Paper 02

edexcel

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 8ST0_02_1806_ER

General introduction

Paper 2 proved to be challenging for some students, though the paper was shown to be less challenging than paper 1, but some achieved high marks. Students still struggled on many questions due to the new style of question, driven by the requirement for the Statistical Enquiry Cycle to be tested, in this 2017 new specification.

Question 1

Both parts of this question tested students on their sampling vocabulary and how it applies to real-life examples. Few managed to score full marks. Students should be comfortable with all vocabulary given in the specification, and they should be able to recognise and produce examples for each of the sampling types.

Question 2

This question is a new style of question criticising use of statistics in real-life publications. This was completed well by students, and almost all were able to provide at least one criticism in context.

Students are reminded that their explanations should be presented as bullet-pointed lists, each written as a clear, specific, and concise sentence. There were several excellent solutions given here.

Question 3

The standard hypothesis test in (a) caused problems for some students. This is likely due to the fact that the data was not separated, though this was regularly seen in the legacy papers. Nonetheless, the modal mark was 9/9, so many students did well here.

Many students gave conclusions to the test which were definite in conclusion, eg 'female frogs are longer, on average, than male frogs'. Students are reminded that hypothesis tests simply provide **evidence** towards a conclusion, and their conclusions should convey this. It should also be pointed out that a **significant** result provides **significant evidence for H_1**, whereas an **insignificant** result provides **insufficient evidence for H_1**. An insignificant result **does not** provide significant evidence for H_0. This should be remembered for all hypothesis tests in the assessments.

In part (b), most students realised the need to use the binomial distribution as a model, and nearly half of students scored full marks here.

Some students did not realise that the answer to part (b) produced evidence for the response in part (c). Some sensible non-statistical answers were often offered here. The required use of information discovered in previous question parts, without direct instruction such as 'with reference to your findings in part (b)', will be seen more regularly in the new assessments than in legacy papers.

Question 4

Parts (a) and (b) were answered well by students, and most managed to at least produce one of the acceptable equations and correctly substitute in $x = 1.2$. Some students calculated the regression line manually, no doubt taking considerably longer than the 3 marks awarded required. Teachers and students are reminded that marks are assigned on the assumption that most calculations will be completed on calculators, and so students should be familiar with all of the relevant functions available on their calculators.

With the expectation to complete calculations using the calculator in the new assessments, it becomes more important for students to show their working in other ways. Many students sensibly crossed out the row for 'UV Ceti' with the missing data, and then crossed out their perceived outlier. If this outlier was incorrect, a mark was awarded for 'incorrect outlier'. Without this clear crossing out, a mark was not awarded. This also affected the A1 mark in part (b) as this was dependent on an acceptable equation seen in (a).

Part (c) was a question on criticising the validity of a statistical conclusion, and there is a requirement for this to be examined in the new specification. Many students gained a mark here. Incorrect answers here tended either to be unclear or vague. Students are reminded that explanations need to be clear, specific, and concise. Again, just a bullet point will suffice.

Part (d) was a context-led question on suggesting improvements to statistical processes. Most students recognised the need to separate star types, and many were able to connect back to parts (a) and (b) and realise that new lines of regression were required. Incorrect answers tended not to be relevant to bivariate data, eg 'use a normal distribution model'. Students should be comfortable with the clear distinction between univariate and bivariate data, and that a univariate statistical technique will not be useful in investigating the relationship between two variables.

Question 5

Part (a) had two standard probability questions which were completed well by nearly all students. Most errors here were due to students not recognising the **total** column on the left and/or **total** row at the top, both of which are non-standard (but this is how this information was presented on the website that published the table).

Part (b) involved constructing a Venn diagram from a two-way table where two columns needed to be combined. About half of students were able to score full marks here. Several students neglected to draw a box around the diagram, leading to some students omitting the fourth frequency $(S' \cap L')$. Some students missed the instruction to include **frequency** data, giving probabilities instead. Students are reminded to read the question carefully, paying particular attention to instructions given in bold text.

Part (c) was a standard conditional probability question. Many students presented an incorrect answer with no working, and as a result scored zero. Students should be reminded to show their method in solutions to questions worth more than one mark.

In part (d), few students scored marks here. Some students did not clearly show that they understood the mathematical consequences of two events being independent. Many others were unable to present their argument in a coherent way. For the B1 mark, a clear statement that these probabilities were not equal was required. Many students began by stating that the probabilities were equal, thus assuming the events independent, and never clearly stated that this was not the case.

Question 6

Part (a) contained two standard calculations for a discrete random variable and about half of students managed to score full marks here. The lack of success was doubtless due to the disagreeable probabilities found in the table. The new specification includes a drive towards realism, and thus will contain fewer neat numbers and probabilities than the legacy specification. Students may need to gain more confidence in handling such 'real-life' data.

Many students calculated the probabilities manually, no doubt taking considerably longer than the 3 marks awarded required. Teachers and students are reminded that marks are assigned on the assumption that most calculations will be completed on calculators. $E(M)$ and $E(M^2)$ can be quickly calculated using the statistics mode on a calculator and the variance can be quickly calculated from these. It is recommended that all students are very familiar with the functions available on their calculators, and it is recommended that teachers focus their lessons using the knowledge of these functions (schools are advised to have a standard calculator recommendation for the whole class to aid teaching of technology).

Part (b) asked students to compare two distributions, though this was a more challenging question than those presented in the legacy specification. For the first B1, students needed to show that they knew to obtain the square root the variance or square the SD before comparing the values. Many students here simply said, eg 'the SD for cars is less' with no evidence that the correct process had occurred. If these values are not given explicitly in the question, students are advised to present them as a numerical comparison.

For the second B1, students needed to show that they understood that the expected value was equal to the mean, which allowed them to compare these values. Many students skipped over this step and did not demonstrate their understanding here. Again, if means are not explicitly given, then a clear numerical comparison should be given.

In the final E1, students needed to demonstrate their understanding of the use of a mean (as an average) and the SD/variance (as a measure of spread). In the new assessments, students will be expected to use their judgement on whether two values are roughly the same. In this case, the measures of spread were very close, and in real life the spread of the two datasets would be considered roughly the same.

Part (c) produced some very well presented and insightful comments, though few students managed to score full marks. Students are reminded that if more than one point is required for an explanation, they should be presented as bullet points with each written as a clear, specific, and concise sentence.

Most incorrect solutions here were due to a lack of clarity (eg 'different dates') or vague non-contextual explanations (eg 'data is biased'). Others mentioned that the sample sizes were different, which will not cause any issues in the real-life context.

Question 7

Part (a) was a standard hypothesis with an additional report added at the end. There is a requirement to test writing and interpreting text aimed at a target audience in the new specification.

This also tested students on multivariate datasets, as students were expected to interpret the table and decide which data was pertinent to the question. This was new for students, and only a few students completed the test completely successfully. Many students made a very good attempt however.

Relatively few students attempted a report at the end of the question. Even fewer students reported Nadia's findings in addition to their own. Students should be reminded to read the full question carefully.

Part (b) was another question suggesting improvements to statistical investigations, a feature of the new specification. This was completed well, with most students gaining at least one mark and resulted in a modal mark of 2/2.

Summary

Based on their performance on this paper, students should:

- read the question carefully and fully before answering the question. In particular, look to see if there is more than one instruction in a question part, and look out for words shown in **bold** type.
- be familiar with all of the functions available on the calculator which can save time in the examination.
- know and understand all **vocabulary** used in the specification, including identifying or producing real-life examples.
- use **bullet points**, each written in clear, specific, and concise sentences for explanation questions.
- remember that not all explanations are statistical but may require some basic general knowledge and understanding in places.
- remember that individual parts are all connected through the context, and previous parts may be required in order to answer a question.
- write conclusions to hypothesis tests in terms of **evidence**, rather than as a definite conclusion.

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