Pearson

Examiner's Report
Principal Examiner Feedback

Summer 2018

Pearson Edexcel GCE
In Statistics (8ST0) Paper 01

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_01_1806_ER

General introduction

Paper 1 proved to be challenging for many students 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

This quick question tested students on their vocabulary. Few managed to score full marks. In the new specification, there is a requirement to test vocabulary thoroughly and therefore all students should be comfortable with such vocabulary and able to gain marks from knowledge of statistical terms.

Question 2

In part (a), few students managed to present the term 'symmetrical', although many students managed to produce a description which sufficiently explained the symmetry.

Part (b) was done well by most students. Most errors here involved dividing by 0.011834 or multiplying by 84.5. Students should remember that **probabilities** are represented by the **area** in this diagram.

In part (c), many students missed the need to now use the binomial distribution as a model. This was sometimes tested in SSO1/MBS1 in the AQA Legacy Statistics qualification but was commonly seen in Edexcel's S2 module. Students should look out for this in future assessments.

Question 3

Most were able to name the sampling technique and its advantages in (a) and (b). Note that context was not required for full marks and hence students who knew their definitions did well here.

Part (c) was a requirement for a simple hypothesis to test for proportion in a binomial distribution. However, few managed to spot that this was the case, no doubt since 'binomial distribution' is nowhere to be seen. In the new qualification, students are expected to be able to recognise a binomial distribution without prompting. Also, sight of the words 'test' and 'proportion' should be sufficient to guide students to the correct hypothesis test. It was fully understood during the standardisation process that students had found difficulty accessing this question part due to a lack of experience of selecting a procedure without a prompt.

Many students gave conclusions which were definite in conclusion, eg 'the proportion of diabetic people under 25 in the population equals 0.61'. 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.

It should also be noted that when asked to "make any necessary assumptions", though it is a good habit to note the assumptions, this will not necessarily gain marks. If assumptions are specifically required to be stated, then marks will be awarded.

In part (d), as the question was entirely in context, explanations were required to be in context also. Some students simply wrote vague explanations such as 'sample may be biased' and therefore lost marks. Students needed to consider the context and tell us **why** the sample might be biased.

Question 4

In part (a), most managed to do the 'maths part' of using a normal distribution to find a probability. However, few attempted to describe the distribution as required, and even fewer attempted to explain why it was chosen. Students need to make sure they have read the full question, as one feature of the new specification is that they may be asked to present multiple points in a single question part.

(b) was a very heavily context-based question that many students struggled with. This style of question features more regularly in the new specification and students are encouraged to try to get into the context rather than viewing the exam as a series of maths-led questions.

Students should look out for small hints given by the question, such as the word '**first**' written in bold, which suggests the first explanation in the mark scheme.

Finally, explanations should be focussed by using the number of marks awarded as a guide. Here, a single mark was available and therefore a single point was required. This should be presented as a clear, specific, and concise sentence or bullet point. The number of lines given for the answer allows for errors and should not be used as a guide for length. Many students wrote a small essay for this part, which was a waste of their exam time. There were some perceptive, correct solutions seen.

In part (c), students struggled due to the slightly different style of question to the legacy papers. In part (a), almost all students decided that a normal distribution was suitable, then one page later when asked to work with a mean of a sample from the distribution, many students seemed to have forgotten that they were using a normal distribution model. In the new specification, more so than the legacy students are expected to be able to use results found in previous parts of a question. If students get stuck on a later part of a question, perhaps looking back over the initial stem and answers to previous parts would be a good place to look for a hint. As the new specification becomes more familiar and there are more past papers, this will doubtless improve.

Part (d) is another context-led question on criticising and improving statistical processes and there is a requirement for this to be examined in the new specification.

Many students missed the instruction to give **two** negative criticisms, and a few missed the instruction to describe **improvements** as well as criticisms. Students should read the **whole question carefully** before answering. Answers should be presented clearly with bullet points to distinguish between points, each written in a clear, specific, and concise sentence.
e.g.

- Temperature at the end of October is likely to be colder than at the start.
  I would rectify this by splitting the data by week instead of month.
- The data does not take global warming into consideration.
  I would change my model to only include data from the last 20 years.

This would have scored full marks, and this is how it is recommended that explanations are presented.

Question 5

Most students successfully answered this part. The most common error in (a) was students stating that the sample sizes were different. A small consideration of how they would approach an investigation into correlation should have confirmed that the data would need to be paired by country first, and sample size can be easily rectified by choosing the smaller of the two sample sizes.

Part (b) is a new style of question for the new specification. Many students were confusing a spreadsheet with a database, and 'query' and 'field' are not terms used in spreadsheet software. However, some students made excellent comments and clearly had a sound understanding of the requirement for knowledge of database and spreadsheet terminology.

Many students completely misunderstood the question, and wrote detailed instructions on how to investigate correlation in a spreadsheet program, e.g. 'plot a scatter graph', 'use the PMCC function' etc. Students need to read the question carefully to see what is required for the marks.

Several students went into too much detail e.g. 'Go to the **insert** ribbon. In the **charts** section, click on the **scatter graph** icon, select the tea and coffee data…' etc. Though this is impressive memorisation of the software, it is not necessary and will not score any bonus points.

Students are encouraged to practise using spreadsheet and database software with raw data in multiple tables, experimenting with manipulating and linking data, and analysing various aspects of the data to see how these are used in real life contexts.

The standard hypothesis test in (c) was done well by most students. However, few managed to describe the assumptions necessary for the test to be valid. Students should ensure that they know all assumptions required for each test, as described in the specification. Since (d) was dependent on this assumption, very few managed to score marks here. Again, this style of follow-on question will be more common in the new assessments.

Part (e) was completed well, with the majority knowing that Spearman's rank correlation coefficient is an alternative to Pearson's PMCC without the assumptions.

In part (f), students were required to read the question very carefully in several places to identify the disparity between kgs and cups, and then use reasoning to explain why this is an issue. Students are reminded that not all explanations are statistical, and that they should rely on their general knowledge and understanding in places. There were several excellent answers here.

Question 6

Again, part (a) was a statistical test that was not precisely described for students. However, this question was done well by most, with the modal mark being 11/11. In such situations, students are encouraged to try to think logically about the data before deciding not to attempt the question. If testing for association between categorical data, there is only one test to use (chi-squared). If considering which hypothesis test we can do with a two-way table, there is only one test (chi-squared).

In this question, three marks could have been gained without even attempting a chi-squared test, including one for identifying the correct data. Students should make sure to highlight appropriate data in large datasets to show that they have understood the question. Nonetheless, a significant number of students simply left this question blank.

Part (b) was another contextual question, this time analysing data as it is presented in real-life. Students were again expected to combine careful inspection of the paper with their understanding of the context. Students found this question challenging, and it was left blank by most students, however this may also possibly be due a time issue.

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.
- know and understand all **vocabulary** used in the specification, including identifying or producing real-life examples.
- use **bullet points** with clear, specific, and concise language for explanation questions.
- try to get into the mindset of the context, rather than viewing the exam as a series of maths-led, routine questions.
- remember that not all explanations are statistical but may require some basic general knowledge and understanding in places.
- 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