Examiners' Report

Principal Examiner Feedback

Summer 2019

Pearson Edexcel GCE Mathematics

In Statistics

Paper 3: Statistics in Practice (9ST03_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 2019
Publications Code 9ST03_01_1906_ER

General introduction

Two aspects of Paper 3 proved to be particularly challenging to candidates. Both were newly examinable features of the 2017 specification. The first was the new style of question designed to test knowledge and application of the Statistical Enquiry Cycle. The second was the requirement to select the appropriate hypothesis test for a given context. Candidates should be given ample opportunity to practise these skill before the exam.

For Statistical Enquiry Cycle questions, it is suggested that candidates note the number of marks available and aim to provide as their response, at least that many clearly-written, bullet-pointed comments, in context, and in full sentences. Many marks were lost on this paper by candidates who wrote responses that were vague in nature or not within the appropriate context.

Additionally, within the examination itself, candidates should be encouraged to use appropriate technology for performing almost all statistical calculations. This should, not least, allow them to spend more exam-time deciding which hypothesis test is appropriate and providing considered, well-written responses to Statistical Enquiry Cycle questions.

For example, for all three hypothesis tests on this paper, it was possible to calculate the test statistic and associated p-value using a graphical calculator. Despite this, the majority of candidates wasted time and sometimes lost accuracy marks by doing their working long-hand. Unless explicitly asked for, it is not required that long-hand working be shown, if numerical answers are correct.

Note also that hypotheses should make reference to population parameters. On this paper, candidates that chose to write their hypotheses in sentences rather than symbols, rarely made the reference to a **population** parameter explicit, and therefore did not gain the mark.

Very few candidates used the p-value to draw a conclusion to their hypothesis tests even when they had written it down. This is an explicitly examinable technique on the new specification; as in Question 2b) on this paper. Candidates performed particularly badly on this question. Furthermore, the critical value approach, while acceptable, is a throw-back to a less technologically advanced age. Candidates are unlikely to encounter it in higher studies or in their place of work. Consideration should be given to increasing the class-time devoted to teaching the p-value approach.

Question 1

Q1 parts a) to d) were generally well-answered. However, markedly fewer candidates understood how to evaluate the median of a discrete distribution than the mode and the mean.

Q1e) Very few candidates gained full marks in this section but most gained at least one mark. The incorrect answers tended to focus on irrelevancies such as the central limit theorem and the sampling technique.

Note that the new specification requires that candidates use the correct statistical language such as “symmetrical”, “skew” and “bell-shaped” in their responses to such questions. Marks were lost through the use of incorrect or unclear language.

Question 2

Q2 The Cohen's d effect size is a topic new to A level statistics. Despite this, many candidates correctly calculated and interpreted Cohen's d. However, very few candidates indeed demonstrated that they knew how to conclude a hypothesis test when the p-value had been provided to them.

Question 3

Q3a) This type of question, involving the interpretation of real-life graphical visualisations, is also new to A level statistics and candidates' lack of practice at this skill was evident. Very few candidates scored full marks. Almost all however, managed to score at least one mark.

Despite the requirements of the new specification, surprisingly few candidates used the correct statistical terminology to describe the overall patterns to be seen in the data, such as "trend", "seasonal" or "cyclical".

Common errors observed were candidates commenting about popularity or sales or price rather than relative search interest; giving a detailed "the line goes up at x days and down at y days" narrative; or giving an answer referencing the weather or seasons, features which were explicitly left off the graph provided.

Q3b) This was generally well-answered but only a small minority of candidates suggested that the frequency of searches would be a useful addition to the data provided.

Question 4

The sections concerning the Poisson distribution were much better answered than those about the Exponential distribution.

Q4a) Pleasingly, most candidates were able to give an appropriate answer, in the correct context, to score at least one mark here. Indeed, several manged to achieve full marks on this challenging question.

Q4b) & c) Were generally well-answered.

Q4d) to f) Around half of candidates recognised and understood the application of the exponential distribution here and performed reasonably well. The others generally scored no marks at all.

Question 5

Q5a) The most common error here was to perform a z-test despite the small sample sizes, and the unknown population standard deviation. Candidates who did select the correct t-test performed very well in this question. However, the hypotheses were generally poorly-stated.

Q5b) This was a new Statistical Enquiry Cycle style of question. Candidates were required to place their statistical results into a report that also considered the limitations of the statistical methods used, and the practical context involved. The majority of candidates made a considered *attempt* at this question and gained at least one mark out of four. However, the lack of practice at such questions was evident and very few indeed scored full marks.

Question 6

Q6a) Most candidates selected the correct hypothesis test for this context. Hypotheses and conclusions were generally acceptable. However, the complicated calculations involved were often not successfully completed long-hand. Candidates who used appropriate technology fared better.

Q6b) Here the majority of candidates could make at least one criticism of the method of data collection. Fewer made reasonable statements regarding the assumptions of the test. The most common error was to criticise the unequal sample sizes.
Q6c) The context of this question was obviously well-understood by candidates and responses were generally more cogent and well-written here than in the other Statistical Enquiry Cycle parts of the paper; with most gaining at least two marks of the four.
Poor responses dismissed the requirements for samples to be representative and unbiased because of the stated time-pressure. Good responses gave imaginative suggestions for adhering to these requirements, despite the time-pressure.

Question 7

Q7a) By candidates that reached this last question with sufficient time remaining, the calculations for the ANOVA were reasonably well performed. This was particularly true if appropriate technology was used.
However, there was generally very poor understanding shown of the meaning of rejecting the null hypothesis for an ANOVA. This was demonstrated both in the phrasing of the hypotheses and in the conclusions drawn. It should be noted that without extensive post-hoc pairwise analysis, the only conclusion that can and should be drawn is that the **two** categories with the most different means, differ from each another, on average. This is substantively different from concluding that one differs from all of the rest.
Despite prompting in the phrasing of the question, the majority of candidates failed to calculate the mean recovery time for each of the times at which the burn occurred.
Q7b) Very few candidates raised the issues associated with a research methodology that utilises secondary data.
Q7c) Most candidates were able to identify a potential blocking factor for the next stage of research.

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