Examiners' Report
Principal Examiner Feedback

Summer 2019

Pearson Edexcel GCE AS Mathematics

In Statistics

Paper 2: Statistical Inference (9ST02_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 9ST02_01_1906_ER

General introduction

The new specification requires candidates to be able to make statistical decisions based upon the given information. Candidates are advised to read **all** of the information given in a question as far too many missed important clues and as a result embarked upon an incorrect method when carrying out their hypothesis tests. The statistical functions on calculators are expected to be used throughout this paper.

Question 1

This question proved to be very accessible to all candidates. Part (a) was generally well answered with the majority of candidates being able to calculate the correct 99% confidence interval for the given data. The most common error in part (a) was to use the value 2.3263 for z from $P(Z<z)=0.99$ rather than using a central area of 0.99 giving the correct z value of 2.5758. In part (b) the majority of candidates were able to comment on the size of the sample and had clearly been trained to recognise that this meant that the Central Limit theorem applied. In part (c) many candidates were able to state the need for either the sample of dolphins to be random or for the mercury content to be independent from dolphin to dolphin. A very common error in part (c) was to state the data was normally distributed after clearly being told in part (b) that this assumption was not necessary.

Question 2

This proved to be one of the more difficult questions for the candidates. The question tests the candidates' knowledge of Spearman's rank correlation coefficient together with a hypothesis test which were clearly familiar to them but many candidates missed the fact that this was a question about correlation. Candidates should be familiar with the term Pearson's product moment correlation coefficient for the PMCC so they are able to realise that the only measure available to use when these conditions are not met is Spearman's rank correlation coefficient.

The candidates who did recognise the need to calculate Spearman's rank correlation coefficient were generally able to correctly rank the data and find the correct value of SRCC to be -0.75. A surprising number of candidates used the differences and the formula to calculate SRCC when it would have been far quicker and simpler to have used the statistics functions on their calculators to find the PMCC of their ranks.

Another common error was with the alternate hypothesis, this was rarely stated correctly with many candidates giving their alternate hypothesis as **an** association rather than the **negative** association suggested by the statement, 'when Orange Tips are relatively abundant, Peacocks are relatively uncommon'.

Unfortunately there was a large number of candidates who chose to use either a non-parametric test, often the Wilcoxon signed rank test, or a paired t-test. These candidates rarely gained any marks for this question. Candidates need to be advised to read every part of the question, considering the relevance of any given information before embarking on answering a question.

Question 3

The standard hypothesis test in part (a) was started well by most candidates. This type of question was common in the legacy SS04 papers. However, few picked up on the need to use the t-distribution due to the sample being small with an unknown population variance. A common error was to use the value of ±1.96 from the normal distribution for the critical value rather than the required t=±2.571.

Candidates should ensure they know all assumptions required for each test, as described in the specification. The question stated 'you may assume that these measurements follow normal distributions' not as an instruction to use the normal distribution but rather as this assumption is a requirement for the use of the t-distribution for small samples with unknown population variance. Candidates who opted for the standard test statistic and critical value approach were generally most successful.

Part (b) proved to be more challenging due to the slightly different style of question to the legacy papers. Most candidates were able to identify the correct t value of 2.26 but were then unsure how to progress to find the critical region. The most common error was to use the normal distribution with $z=1.96$ rather than the t-distribution. Candidates need to be reminded to read all of the instructions given in a question, part (b) asked for answers to be given to an appropriate degree of accuracy so we were looking for answers to 1 decimal place rather than the usual 3sf here.

Question 4

This question was a requirement for a 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, candidates are expected to be able to recognise a binomial distribution without prompting. Also, sight of the word 'test' together with a given percentage should be sufficient to guide candidates to the correct hypothesis test. As the new specification becomes more familiar and there are more past papers, this will doubtless improve. The most common initial error was setting $p=0.51$ in the hypotheses. A majority is more than half so candidates should have been using $p=0.5$ and $p>0.5$ in their hypotheses. Candidates who made use of the first method stated in the mark scheme using the formula given on page 4 of the formula book often scored the most marks. Candidates who chose to use an exact binomial distribution often ran into problems due to the fact that 51% of 1025 is 522.75 rather than an integer value, marks were gained for finding either $P(X \geq 522)$ or $P(X \geq 523)$ and then comparing this p- value with 0.05. This was another example of a question with more than one instruction, candidates were directed to use a hypothesis test and **also** explain, in context, whether this conclusion meant that the energy advocacy's group was definitely wrong. Some candidates did not even attempt this part of the question but it is worth noting that even if their hypothesis test had been carried out incorrectly candidates could still have gained marks at the end for relevant explanations. This part of the question was designed to test whether candidates could pick up on the fact the poll was not random and was only being conducted by telephone or that they could recognise that 1025 was a very small sample compared to the population of USA. Candidates should also be familiar with another important feature of a hypothesis test that whilst one particular sample may well fail to provide evidence to support a claim, further samples could give different results.

Question 5

Part (a) should have been the source of some easy marks for candidates, the statement 'scores are unlikely to be normally distributed' should have lead them to conclude a distribution-free test was required and as the data was unpaired it should therefore have to be a Wilcoxon rank sum test. This was another example of the new style of question. In the new qualification, candidates are expected to be able to recognise when to use a distribution-free test without prompting. As the new specification becomes more familiar and there are more past papers, this will doubtless improve. The most common error was the wording of hypotheses. When the hypotheses are given in words we need to see population median rather than just median. Acceptable wording for the null hypothesis was 'no difference in population median', 'samples from identical populations' or '$\eta_A = \eta_B$' but **not** $\eta_d = 0$ since differences were not required. For the alternate hypothesis acceptable wording was 'population median score for B < population median for A' or '$\eta_B < \eta_A$'. Those candidates who recognised the need to carry out a Wilcoxon Rank Sum test were generally able to rank the data as a whole correctly including the tied ranks and went on to calculate the total ranks for each group, however, a surprising number of candidates then used the wrong formula for U despite the correct version being given in the formula book. Candidates should be given plenty of practice at this type of question, it was disappointing to see a large number of candidates losing the final conclusion mark, when ts>cv then H_0 should **not** be rejected for this type of test.

There were 3 marks available in part (b) which should have suggested the need to give a fully explained answer. The first mark was for stating that it was reasonable to assume the two groups were independent followed by a reason given in full context. Candidates should be encouraged to read all of the information given in part (a) when making their decision. The question clearly stated a random sample of air traffic controllers was taken and that these controllers were randomly assigned to a version of the test. Many candidates chose to ignore this information and instead went into long explanations about the stress involved in the job of an air traffic controller or focusing on the controllers being unable to talk to each other rather than considering the given facts.

Question 6

This proved to be one of the more accessible questions for the candidates. The question tests the candidates' knowledge of the Poisson distribution alongside hypothesis tests which were clearly familiar to them but their answers often lacked the required detail in part (c).

Part (a) was generally well answered. The most common error was to calculate $P(X= 5)$ rather than $P(X\geq5)$ or just 1 minus the sum of the other probabilities.

In part (b) we were looking to see if the data could be modelled by a Poisson distribution. There were many fully correct solutions to this part however common errors included omitting to combine the final three groups due to the **expected** frequencies being less than 5 and incorrect degrees of freedom. The mean had been calculated from the data in part (a) so the degrees of freedom should have been 4-2=2. Candidates are encouraged to write down their degrees of freedom, to enable follow through marks to be awarded for an incorrect critical value using the chi-squared distribution with their degrees of freedom. Candidates are also advised to show some sort of method for their attempt at $\sum\frac{(O-E)^2}{E}$ incorrect final answers from a calculator cannot score method marks unless a method has been shown, showing at least two of their substitutions would have sufficed. The final conclusion to any hypothesis test should always be written in the context of the question.

In part (c) many candidates were unable to make the connection between the conclusion they had just drawn in part (b) and what this told them about the arrival pattern of customers. As part (b) had given evidence to suggest the Poisson distribution was **not** suitable for the number of customers visiting the ATM during the evening candidates should have been referring to the properties of the Poisson distribution not being met in the context of the question. Suggested answers included customers do not appear to be arriving independently of each other or that customers did not appear to be arriving at a constant average rate. Alternatively they could have referred to the largest contributions stating the first and last categories had far more observed values than expected suggesting there were more quiet and busy times than a constant average rate would have suggested. A common error in this part of the question was to refer to the time of the evening affecting the use of the ATM usually stating more people used the ATM earlier in the evening when this information had not been given in the question.

Question 7

In part (a) there was widespread confusion on whether to use a one-factor or two-factor ANOVA table. It was accessible in that almost all candidates did attempt either one of these and it was heartening to see so many candidates give this question a really good go. Candidates need to be trained to spot the clues in this type of question when making their decision. The given table had both row and column totals and the initial stem referred to the 3 energy drinks and the 6 volunteers, these clues should have been enough to suggest a two-factor ANOVA. Those candidates who made use of the statistical functions on their calculators were often the most successful and were then able to use the gained time to concentrate on the later explanations in parts (b), (c) and (d). It is worth training candidates to use these calculator functions but to also encourage them to carry out a very simple check of their degrees of freedom and to also show the divisions required to give their mean sum values and final F-ratio value just in case their initial data entry was incorrect.

Common errors using the formulae were mistakes in calculating the sum of squares for rows and/or columns; mistakes in putting the degrees of freedom in either the wrong order or putting the wrong degrees of freedom completely. Possibly an area here which requires greater practice when this topic is initially taught as an error here led to an incorrect test statistic and also an incorrect critical value.

In part (b) the problems usually arose from an incorrectly completed part (a) and thus no differences or rather incorrect differences found. Many left this blank or gave responses referring to the subjects. This is another example of this new style of question requiring candidates to look back at the original information to help with later parts. Part (a) referred to the energy **drinks** differing so answers to part (b) should also have been referring to the drinks.

Part (c) was highly variable and in many cases independent from earlier answers to this question. Randomised Block was often quoted with little response further than this, though common errors were ‘repeated measures’ or ‘double blind’ or even ‘completely randomised block’. Candidates should be encouraged to look at the number of marks available, 3 marks should have suggested more than one advantage was required. Candidates should also be discouraged from simply writing eliminates bias for these sorts of questions.

In part (d) it was clear that many candidates were not clear on what a ‘blocking factor’ was. Of course those candidates who had chosen to carry out a one-factor ANOVA test in part (a) really had little idea of the blocking factor. Many candidates concluded that the blocking factor had been ineffective despite having evidence to the contrary from their earlier test in part (a), quoting that different athletes would perform at different levels. Attention should be drawn to the request for numerical justification in part (d) this should have been a clear indication that some numbers were required either from their original ANOVA table or by carrying out a new one-factor ANOVA test and comparing their two conclusions. The asked for numerical justification was thin on the ground often losing candidates two marks for otherwise very strong answers to this question.

Summary

Based on their performance on this paper, candidates 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.
- ensure they can recognise which hypothesis test to use in given situations and **read** the given information carefully to help decide which test to use.
- make greater use of the statistical functions on their calculators.
- know and understand all **vocabulary** used in the specification, including how to analyse data given in a spreadsheet or database.
- use **bullet points** with clear, specific, and concise language for explanation questions.
- write conclusions to hypothesis tests in terms of **evidence**, rather than as a definite conclusion and always in the context of the question.

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