Pearson
Edexcel

Examiners' Report
Principal Examiner Feedback

October 2020

Pearson Edexcel GCE
In Statistics (9ST0_02)
Paper 2: Statistical Inference

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

October 2020
Publications Code 9ST0_02_2010_ER

General introduction

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. The new specification requires candidates to be able to make a decision about which hypothesis test to use based upon the given information.

This was the area that proved to be the most challenging to candidates. Candidates should be given ample opportunity to practise these skills before the exam.

Additionally, within the examination itself, candidates should be encouraged to use appropriate technology for performing almost all statistical calculations. 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.

Question 1

This question proved to be very accessible to almost all candidates. Part (a) was generally well answered with the majority of candidates being able to calculate the correct 95% confidence interval for the given data. The most common error in part (a) was to use the value 1.6449 rather than the required value of 1.96 for z. However, in part (b) markedly fewer candidates actually compared the given mean of 5.8kg with their confidence interval before making their comment. In part (c) Very few candidates gained full marks but most gained at least one mark. Candidates should be encouraged to give full context in their answers when referring to assumptions. Poor responses simply stated random sample and normal rather than referring to the assumption that the **babies** were a random sample and their **weights** were normally distributed.

Question 2

This standard hypothesis test for a proportion was completed fully by the vast majority of candidates and proved to be an easy source of marks. There were a variety of methods used, p-value using the exact binomial distribution, normal approximation using the formula given in the formula booklet and the critical region approach were all

successful and acceptable methods. However, 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. It was pleasing to see so many candidates giving full context in their final conclusion.

Question 3

In part (a) many candidates chose to compare the given t-value with t_{24} rather than comparing the given p-value with 0.05 both of these methods were acceptable. Even though the calculator output specifically mentioned a 1-sample test a surprising number of candidates mistakenly used n-2 to calculate the degrees of freedom to be 23.

Part (b) was poorly attempted, often totally blank. The required definition for a Type I error is rejecting H_0 when H_0 is true. Only a very small minority of candidates were able to state P(Type I error) as 0.05

Part (c) the most common correct answer was the fact that the children were a random sample therefore the conclusion was reliable.

Question 4

Part (a) should have been a source of easy marks for the well-prepared candidate. The data was given in a **two-way table** together with mention of testing for no **association** between region of the world and sex of player. This information together with reference to **combining** regions should have pointed candidates towards analysing a contingency table. When a question refers to giving numerical justification candidates should be reminded of the need to include calculations or at the very least some numbers as part of their solution. Reference to the expected frequencies being less than 5 was required in part (a).

Part (b) was generally completed well with many candidates making good use of the statistical functions on their calculators to obtain the relevant contributions. However, they are also advised to show some sort of method for their attempt at $\sum\frac{(O-E)^2}{E}$ as incorrect final answers from a calculator cannot score method marks unless some sort of 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. A small number of candidates decided to carry out a hypothesis test for correlation.

Part (c) asked candidates to describe, with numerical justification, the nature of any association found between region of the world and sex of player. Candidates should be reminded of the need to include either calculations or reference to values from their contributions when asked for numerical justification. They could have referred to the largest contribution to the association (4.08) being from women and Eastern Europe or to the fact that there were more women observed (19) than would be expected (12) from Eastern Europe to earn at least $1 million in 2018.

Question 5

Attention should always be drawn to any part of a question in **bold** type. The required average for all English state schools had been given at the bottom of the table in bold type. Part (a) proved to be more challenging due to the slightly different style of question to the legacy papers. Part of the solution to a hypothesis test had been given, candidates were expected to recognise that this was part of the workings from a sign test. Very few candidates correctly stated the relevant hypotheses for Davinder's test. The most common error was stating $H_0: \eta = 0$ and $H_1: \eta > 0$ rather than the required $H_0: \eta = -0.02$ and $H_1: \eta > -0.02$ When the hypotheses are given in words we need to see population median rather than just median.

Part (b) proved to be more accessible, aided by the specific instruction to carry out a Wilcoxon signed-rank test. It was encouraging to see the vast majority of candidates who reached a conclusion using full context, guided by the wording of the question.

In part (c) attention should be drawn to the number of marks available, two marks suggests two pieces of information are required. This type of question was common in the legacy SS03 papers. Candidates should ensure they know all assumptions required for each test, as described in the specification. The assumption required to carry out a Wilcoxon signed-rank test is that the distribution is symmetrical. If this assumption has been met then a Wilcoxon signed-rank test is preferable to a sign test as it is a more powerful tests as it takes account of the sizes of the differences rather than just the signs.

Question 6

Part (a) proved to be a very accessible question, candidates were generally able to describe a double blind trial and appreciate that if patients were assigned either a formula based meal or a conventional meal they and the people assigning the meals would be aware which diet they were following so this could not be a double blind trial.

In part (b) most candidates who attempted this question selected the correct hypothesis test for this context. Hypotheses and conclusions were generally acceptable. A common error was to use the value of ±1.6449 rather than ±1.96 when the normal distribution was used for the critical value.

In part (c) 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.

Part (d) was generally left blank. Here we were looking for some reference to either the patients sticking to the diets or that the sample variances could be used in place of the unknown population variances due to the large sample sizes.

In part (e) candidates who had been able to correctly carry out their hypothesis test in part (b) were generally able to also complete this part correctly. The biggest source of confusion was in the hypotheses with many candidates being unsure what to do with the 4.

Question 7

Part (a) proved to be a source of very few marks with candidates clearly not prepared for being given summarised data. Very few candidates were able to make any progress with the summarised data for this ANOVA question. Although candidates are encouraged to make use of appropriate technology there could also be occasions when they may need to refer to the formulae given in the formula booklet. This question was such an occasion. There was also very poor understanding of the required format for the hypotheses prior to carrying out an ANOVA test.

$H_0: \mu_A = \mu_B = \mu_C = \mu_D$ $H_1: At\ least\ two\ of\ \mu_A, \mu_B, \mu_C, \mu_D\ differ$

Part (b) was generally left blank. Candidates should be encouraged to consider the assumptions required for an ANOVA test before embarking on such a question – tyre lives normally distributed with a common variance. The given diagrams were roughly symmetrical for each brand of tyre suggesting the tyre lives were normally distributed for each tyre brand. Alternatively, candidates could have commented on the spread of the data, either stating the five brands appeared to have a similar spread of data so the test was valid or deciding that brand B had a greater spread so the test was not valid. There were two parts to this question, candidates should be reminded to read **all** parts of a question and ensure their answer covers all the required points, in this case, considering the information in Figure 7 **and** the method Daniel used to collect the data.
Part (ci) was generally well attempted with those candidates choosing to illustrate their answer as a two-way table being most successful as they could then see the need to record data for each tyre brand with both front and rear wheel tyres.
Part (cii) should have been a straightforward mark, following reference to a randomised block design. Randomised block designs are analysed using a 2-factor ANOVA. Two common incorrect answers were the Wilcoxon Signed Rank test and a z-test.