Examiners' Report
Principal Examiner Feedback

January 2023

Pearson Edexcel International Advanced Level
In Statistics S3 (WST03) Paper 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

January 2023
Publications Code WST03_01_ER_2301

General

Candidates were generally well prepared for the demands of this paper with many strong performances seen. Q1 was the most discriminating on the paper followed by Q5 and Q7. Questions involving the chi-squared test remain a strong topic for the candidates at all levels. Candidates should be advised to comment in the context of the question not only when completing a hypothesis test but also when referring to assumptions required to carry out the tests.

Report on Individual Questions

Question 1

This question was seemed to be liked by most students with very few not attempting all the parts. However, it was a question where many students did not score well so there are still a few issues that are worth emphasising.

In part (a)(i) the 'show that' the mean of *x* is 1008.47 did not always make it clear that $\bar{x} = 1008.47$. Incorrect labelling included labelling 1008.47 as $\bar{y}$ while others labelled the value of $\bar{y}$ as $\bar{x}$ while others didn't label anything. Students need to ensure they label the values they have found accurately if they wish to gain the marks.

In part (ii) there were many students who still didn't realise that $\mathrm{Var}(X + a) = \mathrm{Var}(X)$ when '*a*' is a constant with many students adding on 1000 to obtain 1064 for their variance. Others attempted to find $\sum x^2$ from $\sum y^2$ and $\sum y$ with varying degrees of success.

In parts (b) (c) (d) the main issue that arose was a student who did not recognise the difference between the observed sample mean of *x*, 1008.47 and the null value of the mean, 1010. In many cases, the values were reversed or 1008.47 was used for both. There was also much conceptual confusion between the confidence interval for the mean and the critical region for $\bar{X}$.

Other students didn't answer the question as stated but calculated a z-score which they compared with 1.96 or found a significance probability which they compared with 0.025. Although these methods give the correct conclusion to a hypothesis test, they have not followed instructions and subsequently lost credit.

There is a significant weakness in many students' understanding of Critical Regions. Most were able to find appropriate limits but then gave a "confidence interval" type answer. Special Cases in the scheme allowed some credit for these students but the problem remains.

Part (e) was well answered with the majority of students well-rehearsed in answering this question.

Question 2

This question was generally very well answered. There were few errors in the calculation of the coefficient in part (a), and the only issues with the hypothesis test in part (b) were the occasional incorrect critical value and the failure to interpret the result properly in the context of the question. The same could be said for the test in part (d), which was about finishing time rather than finishing position, and inaccurate comments in this respect were seen from time to time. Part (c) was almost universally answered correctly.

Question 3

This question was accessible to many students with the majority scoring full marks. The vast majority of students were able to calculate both expected values accurately. In part (b) most students stated the hypotheses correctly. The only errors made were:

- giving the null and alternative hypotheses the wrong way round
- failing to contextualise the hypotheses and conclusion, often missing the key words 'making a claim' and 'age'.

Question 4

This was another accessible question. The most common error was not to include the specific binomial that they were testing for in the hypotheses and using the wrong degrees of freedom for part (a). The expected values were generally calculated correctly with only a few using a discrete uniform distribution. Mostly correct calculations and conclusions in context followed. The vast majority of students were able to show that the probability that a child is a girl is 0.45.

In part(c) again the main error was regarding the degrees of freedom. However, there were also many students who basically repeated the test from part (a), failing to appreciate the loss of one degree of freedom for the estimated parameter.

There were some very convoluted conclusions e.g. in part (a), after correctly deciding to reject H_0, "there is insufficient evidence to accept the Binomial distribution". There are still too many students who do not know how to give a correct statistical conclusion to a hypothesis test.

Question 5

Part (a) was a good source of marks with most students correctly stating the hypotheses in terms of μ. Some students lost marks due to failing to provide the critical value to at least 4 decimal places, often stating $z = 2.32$ or 2.33 instead of 2.3263 or better. Most students made the correct decision to reject H_0, but students still don't provide a contextual conclusion.

In part (b) students who failed to realise the need to calculate the expected profit per plant for each type of fertiliser hardly earned marks in this part of the question. Very few calculated the profit per plant and go on to conclude that Claire should use fertiliser *B*. A very small proportion of students calculated the results for 90 plants and made the correct comparison leading to the right conclusion. The most common error was to compare the profit for the 50 plants using fertilizer *A* with the profit for the 40 plants using fertilizer *B*.

Overall, part (b) proved to be a very challenging question even to those students who had produced an exemplary solution in part (a)

Question 6

The confidence interval in part (a) was a straightforward 4 marks for most students. In part (b) many missed that the distribution of the weights was given as Normal meaning that the CLT was not needed. Part (c) was a common sense question answered well by most. Part (d) caused confusion as many students did not realise that they were dealing with the distribution of the sample mean with resulting errors in the calculation of the correct variance.

Many students are too anxious to standardise without first showing the variable which is being standardised; hence *y* was seen too often instead of $\bar{y}$.

Question 7

This question proved to be a good discriminator with the most able students able to gain full marks. At the other end of the spectrum the students continue to make the errors in finding the mean and variance.

In part (a) most students found the mean correctly, but a common error for the variance was to use $Q = 3P$ rather than $Q = P_1 + P_2 + P_3$ resulting in a variance of 3600 instead of 1200.

Part (b) saw oft-repeated appeals to the CLT or other aspects of the normal distribution in a minority of cases, but the majority of candidates were able to give the correct assumption in context.

Again in part (c) the problem was the inability to find the correct variance, but there were also a number of students who were unable to find the correct mean.

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