Examiners' Report

Principal Examiner Feedback

October 2021

Pearson Edexcel International A Level

Mathematics in Statistics S3 (WST03)

Paper : WST03/01

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 candidates at: www.pearson.com/uk

October 2021
Publications Code WST03_01_2110_ER

General

Candidates were generally well prepared for the demands of this paper with many strong performances seen. Q5 and Q7 were the most discriminating on the paper. The calculation of Spearman's rank correlation coefficient continues to be well attempted. Candidates should take care when completing a hypothesis test to use correct notation when defining their hypotheses and 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 tests.

Report on individual questions

Question 1

This question was answered well by the vast majority of candidates and provided an accessible start to the paper. A few candidates however lost the first mark as they stated their hypotheses in words or in other letters rather than μ ($\bar{x}$ being the most common). A common error seen was that candidates used 2.5 as the standard deviation rather than $\frac{2.5}{\sqrt{80}}$, obviously failing to realise that the test statistic in a test for the population mean for μ requires $Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$. The vast majority of candidates were able to identify the correct critical value from the tables and many gave a correct conclusion in context.

Question 2

The stating of the hypotheses in this question seemed to cause some candidates problem and a variety of different approaches were seen. The common error was to define the hypotheses the wrong way round with many of these candidates thinking that H_o should be worded to imply that potassium would have an effect on the quality of apples. The majority of candidates were able to score the next two marks as many were able to calculate the expected frequencies needed. Many candidates were then able calculate a correct χ^2 value but some candidates combined grades *D* and *E* presumably because the observed value for grade *E* was less than 5. The vast majority of candidates were able to state the degrees of freedom and the correct χ^2 value from tables. As this mark was a follow through mark, those candidates that had incorrect degrees of freedom gave a correct χ^2 value from tables and so scored this mark. The conclusions given were varied and the most successful responses referred to Andy's belief. Generally when this mark was not awarded candidates lacked the contextual comment needed.

Question 3

Part (a) provided a successful start to this question for many candidates as they were able to score full marks for obtaining a correct value for the coefficient. When marks were lost, this generally occurred when candidates ranked the jam's price/taste in alphabetical order, but these candidates were still able to score the two other M marks.

In part (b) the setting up of the hypotheses again caused candidates issues, with many losing this mark as they wrote the hypotheses in words. For those candidates that did use correct notation some incorrectly thought that a one-tailed test was required. Most candidates were able to identify the correct critical value for a test at the 5% level of significance. Conclusions here were generally given in context.

In part (c) the vast majority of candidates scored full marks as they were able to use the given values to correctly calculate the value of the product moment correlation coefficient.

In part (d) the setting up of the hypotheses caused less issues than in part (b) but again those candidates that wrote the hypotheses in words lost this mark. Again most candidates were able to identify the correct critical value for a test at the 5% level of significance. Again conclusions here were generally given in context.

In part (e) very few candidates scored this mark. Many identified the incorrect test and for those that did the reason given was often incorrect or irrelevant. Only the very best candidates gave a reason that suggested that a joint normal distribution was unlikely.

Question 4

Part (a) of this question rarely scored full marks. Some candidates did not adequately describe the four lists from which they were selecting the households and some did not include the use of random numbers to select an appropriate number of households from each area. However, many candidates were able to calculate the four sample sizes and so SC B1 was often awarded.

In part (b) the majority of candidates were able to correctly calculate the two expected frequencies needed for the required age group.

In part (c) too many candidates did the full calculation and failed to realise that the $\sum \frac{(O-E)^2}{E}$ for the other four classes was given in the question. However in many of these cases the correct χ^2 was calculated. A few candidates that did the full calculation then went on to add this to the given value and so lost the marks that were available here. Some candidates simply used $\chi^2 = 4.657$, which was given in the question (with no additional calculation), for the remainder of the question. The degrees of freedom stated were generally correct and candidates were able to give a correct χ^2 value from tables. Again, as with other questions that required a contextual conclusion, too many candidates lost a mark as their answer lacked the required wording for this mark to be awarded.

Question 5

In part (a) many candidates were able to correctly calculate the 99% confidence interval required. Some candidates lost marks as they used an incorrect z value and some lost marks as they used $\bar{x} = 3$ (which was the stated weight printed on the bags of flour) rather than 2.977 (which was the mean weight of the sample).

Part (b) generally caused candidates difficulty and only the better candidates were able to give an explanation that linked the confidence interval with the stated weight. Those candidates in part (a) who had used $\bar{x} = 3$ should have realised that they had done something wrong in part (a) and gone back and corrected their mistake, but many failed to do this.

In part (c) many candidates were able to set up and solve the correct inequality needed. However, some candidates used an incorrect z value and so lost the two A marks. Common errors seen for the standard error included $\sqrt{\frac{0.015}{n}}$ or $\frac{0.015}{n}$. A few candidates lost the final A mark as they gave an answer of 55 or 2.711 (which came from $\sqrt{7.35}$).

Question 6

Part (a) was answered well by the vast majority of candidates and virtually all candidates scored 1 mark as they could write down the unbiased estimate of the mean. Candidates that lost marks in this part did so because they did not calculate an unbiased estimate of the variance.

In part (b) the defining of the hypotheses again caused candidates some difficulty. When using letters to identify those that exercise regularly and those that don't, these need to be explicitly defined in the solution. Too many candidates had hypotheses like $H_0 : \mu_x = \mu_y$ without actually defining what x and y represented and so the B mark was not awarded. A common error seen in the calculation of the test statistic was the use of $\sqrt{\frac{16.693^2}{50} + \frac{29.6^2}{40}}$. The vast majority of candidates were able to identify the correct critical value from the tables and many gave a correct conclusion in context.

In part (c) most candidates were able to write a comment that referred to the CLT, however very few candidates made the connection that this allowed you to assume that **both** groups were normally distributed.

Part (d) was not answered well by the majority of candidates. Those that did score marks usually did so because they realised that they needed to assume that the population variance = sample variance. Very few candidates realised that the groups or males needed to be independent and generally wrote a comment referring to the need for the samples to be taken randomly. As this was given in the question no credit was given to these types of answers.

Question 7

In part (a) many candidates were able to find the required expected value and the variance needed and so the first 2 marks were often awarded. Too many candidates failed to realise that they needed to consider both parts and M0 M1 A0 was a common mark distribution here.

Part (b) caused candidates some difficulty. A common error for the distribution was to use $\overline{B} \sim \mathrm{N}(1.96n, 0.003n)$. Again a few candidates used an incorrect z value (usually 2.5758 or 3.0902). Some candidates after setting up the correct inequality went on to get an incorrect answer usually due to some poor algebra.

Part (c) discriminated well from those that could and those that could not. Some very neat and concise solutions were seen. Generally, candidates were able to find a correct mean for M but finding the variance of M was more problematic. Many were able to realise that they needed to find $4T - 3M$ or equivalent and again were able to find the correct mean, but as before candidates struggled to find the correct variance. A few candidates lost the final mark as gave answers to two significant figures (0.087).

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