Examiners' Report
Principal Examiner Feedback

Summer 2024

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 candidates at: www.pearson.com/uk

General

Overall, this paper allowed all candidates to demonstrate their ability and knowledge of the WST03 specification. It was pleasing to see some very strong performances on questions and clear structure to some solutions. However, on a number of questions, candidates did seem to misunderstand what was being asked of them and proceeded down a more anticipated route, as opposed to appreciating the subtle differences between these questions on this paper and maybe ones they have seen on previous past papers.

Question 1

This was a short opening question to the paper on systematic sampling in the context of selecting employees from an alphabetical list.

In part (a), many candidates were able to say that every 50th should be selected, but although they frequently said to select a random number, many failed to make it clear that this was the starting point or to make it clear that the employee with this number should be chosen. There was the usual confusion by a number of candidates with this sampling technique of dividing into 8 groups of 50 and selecting the same position in each group.

In part (b), candidates rarely engaged with the actual context, so it was always pleasing to see responses such as "two employees with the same surname are less likely to be chosen" rather than "may be biased". Recognising general issues were broadly condoned on this occasion, but candidates should be reminded to relate their knowledge of such techniques to the actual context.

In part (c), it was surprising that the answer of 0 did not appear very often. Indeed, many tried to calculate $^{50}C_8$ or $^{50}P_8$.

Question 2

This question tested the topic of Spearman's rank and testing for positive correlation. Most completed this well, although candidates should realise that by showing their working fully, they stand to minimise the loss of marks due to small slips.

In part (a), candidates were usually successful in achieving the coefficient of 0.821 (3sf). However, greater care should be taken when ordering data. Some candidates did get confused with the table given in the question and just assigned values of 1 to 7 for A to G. Provided candidates show sufficient steps in their working, they should still be able to score two marks for finding the differences, squaring and using the formula, however, the rankings were often written around the table and sometimes it was not very easy to see. Sometimes the rankings were not written at all, despite the instruction to "Show your working clearly". Even more surprising was when candidates used a rank of 1 for the lowest priced kettle despite the instruction in the question to use a rank of 1 for the highest.

In part (b), candidates were usually able to form the correct null and alternative hypotheses. However, some did not appreciate that being able to rank kettles in order of decreasing price meant that a hypothesis test for positive correlation should be formed. Others just formed a two-tailed test thinking that a positive or negative correlation would mean that you could generally rank kettles correctly or considered decreasing price as leading to a negative correlation. Others were influenced by their errors in part (a). Despite this, marks were still available for finding the correct critical value and making a comparison as to whether the null hypothesis should be rejected or not. A few drew the opposite conclusion for their combination of r and the critical value, but the contextual conclusions included the required words in the vast majority of cases.

Part (c) was often brief in detail and so some of the responses were vague. Many candidates had the right ideas, but it was clear that many did not appreciate the difference between product moment and Spearman's rank correlation coefficients. Indeed, had candidates taken into account that Aarush's data was already ranked meant that the product moment correlation coefficient would not have been suitable. Some candidates just stated contradictory statements such as "ranks are non-linear" which suggested some vague recollection of what the underlying theory was, but they were unable to apply it correctly.

Part (d) was generally well answered with most recognising that A and D would have tied ranks. The most common error was to say that the ranks for both would be changed to 1 instead of 1.5, and a few either neglected to say which two kettles would be tied or gave the wrong pair. A number of candidates used "it" or "they" which it would be clearer if candidates specifically defined what they were to make sure that sufficient detail is given in their responses.

Question 3

This question tested knowledge of confidence intervals with a final part relating to the binomial distribution. Seemingly, for many it was clear that they are very accurate at finding confidence intervals, but still struggle with interpreting what they actually represent.

In part (a), most candidates were able to confidently state the hypotheses in terms of μ, however many missed explicitly stating that it was a 5% significance level. It was surprising just how many candidates attempted to carry out a full hypothesis test in part (a), which often included finding values required for part (b). A lot of time was spent by candidates on this incorrect approach, so it is always worth looking at the number of marks available. For a three-mark question which required hypotheses to be stated, a significance level to be stated and a conclusion, there was never going to be any more method required other than to look at whether 328 was within the confidence interval or not. Many seemed to conclude that because 328 was within the confidence interval that we should reject the null hypothesis and that there was evidence that the mean was not 328, possibly seeing the interval as a rejection region.

In part (b), many of the candidates were able to score all six marks, however a lack of accuracy with the z values did mean that some were unable to find accurate enough answers to their confidence interval. It was also common for candidates to round their value for σ to 6, rather than use a more accurate value of e.g. 5.9987 which would have given them a more accurate answer. Some candidates did not correctly find the mean of the random sample of 150 bottles and instead used 328 to find σ.

Part (c) was a more mixed response, with candidates stating the wrong values for the binomial distribution or incorrect inequalities for the values that they were finding the probabilities for. A pleasing number, however, did proceed to find the correct answer.

Question 4

This question tested the topics of goodness of fit and contingency tables. Typically candidates are very strong with this topic and they can often be a good source of marks. However, many misunderstood what was being asked and tested for association between age and favourite ice cream flavour in part (a) instead. Part (b) and part (c) should really have made candidates consider their approach in part (a) as they should have realised that they were calculating the same values as they had done previously, but some proceeded to just state the same values again without any further consideration.

In part (a), the majority of candidates included the ratio in their hypotheses, however some candidates simply referred to 'the proportion is correct' rather than being more specific with the values given. Others were far too brief and just stated hypotheses such as "the manager is correct" against "the manager is not correct". As a model with a ratio had been provided, this should be given in the hypotheses. Others copied their null hypothesis for their alternative hypothesis but forgot to include the word "not" to create the opposite scenario. Many candidates created more work than necessary for the rest of part (a), even if the ratio had been stated, pursuing to find incorrect expected values which were not based on the ratio. As a consequence of candidates carrying out this incorrect hypothesis test, it was also common for candidates to have 6 as their degrees of freedom rather than 3 (i.e. the answer to part (c)). Therefore the only marks available for these candidates were the value from chi-squared distribution for 6 degrees of freedom and a correct comparison using their calculated chi-squared value and their critical value. For the final mark in part (a), candidates need to ensure that they give their final statements in context and refer to the flavours of ice cream (or reference to the manager's belief was condoned here).

Part (b) was answered well with many candidates able to score full marks for this part, even if they had lost marks in part (a).

Candidates also seemed confident with part (c), so most candidates scored at least 3 marks on this question, although many will feel that, had they realised their misconception at the start, the marks scored would have been considerably higher.

Question 5

This question was based on a hypothesis test for the difference between the means of two Normal distributions with variances known, as well as calculating a sampling variance from given information. There was a wide range of marks, although achieving full marks was extremely rare.

In part (a), many candidates were confused in their hypotheses, getting the subtraction the wrong way around or stating that the means were equal in H_0. Some got the part-time and full-time employees the wrong way round and there was the contextual misunderstanding that being faster actually meant that times would be less. Some incorrectly had the difference between the two means set equal to 0 for the null hypothesis, whilst the "1" then appeared in the alternative hypothesis. Most were able to find the standard error accurately, but in many cases their standardisation was incorrect due to the order on the numerator. Candidates were usually able to score for a correct critical value, but often no further marks could be scored. Those who did proceed correctly, pleasingly achieved a z value of 1.24 and were usually very successful and often able to contextualise their conclusion fully using their critical value.

In part (b) many did not clearly reason about **both** groups and those that stated independence as a reason were referring to the groups of employees rather than the individual results. Others stated about samples being large but did not provide further explanation relating to the Central Limit Theorem. More detail would have benefited the responses seen as clearly candidates were generally on the right train of though.

Part (c) was typically only partially attempted, with most candidates able to calculate $\bar{a}$. In many cases $\sum a^2$ was incorrectly calculated, and candidates either stopped at this point or if they did proceed further, the actual method for s^2 was not correct. A number of candidates would have benefited from setting out their method more clearly, considering the formula to find the sample variance and then proceeding to find the required elements. Despite the difficulties for a significant number, there were many candidates however that were able to complete a full and thorough method which was clear to follow.

Question 6

This question was on the distribution of linear combinations of independent Normal random variables. It was clear that candidates were confident with most of this topic, although part (c), provided challenge for many so it was rarely answered correctly by most.

In part (a), it was typical for candidates to score full marks. They were confident in finding the mean and the variance which was required for the scenario and mostly candidates were able to standardise correctly. Occasionally candidates would attempt to work with just C rather than $3C$ and the most common error in finding the variance was to attempt $9\text{Var}(C)$. Some candidates forgot to use the standard deviation rather than the variance. Some candidates subtracted their answer from 1 at the end which cost them the final mark.

Part (b) was usually answered well, although the most common mistake was failing to double their probability for $P((R-R)>0.05)$ which cost them the final two marks. Candidates typically found the correct value of the mean for the distribution, although those who made the error in calculating the variance in part (a) often made the same error in part (b). Overall, however, these first two parts were a good source of marks for a number of candidates.

Part (c) was problematic for many candidates, however, so a correct answer was rare. The issue possibly was in the subtle difference from the January 2024 paper where one of the variables described in the final part was the same variable which enabled "terms to collected" in establishing a new variable. However, in this case the variables G and T were independent so the tray which was part of the variable G was not the same tray as the tray described in $G<2T+20$. Therefore, many incorrectly rearranged the inequality $10R+T<2T-20$ to establish a variable $10R-T$ whereas T should have been considered as T_1 and T_2. Even with the incorrect method, the mean for the combined distribution would still be correct, but it did impact on the variance. Therefore, only a maximum of three marks could be scored by those who had formed this incorrect distribution. Other common errors appeared when standardising such as forgetting to square root the variance, or simply miscopying earlier values.

Question 7

This question tested understanding of the distribution of the sample mean for a continuous uniform distribution and then using a confidence interval to find the size of the sample. Whilst candidates scored well in part (a), unfortunately part (b) was rarely answered correctly; it was extremely uncommon for candidates to score full marks.

In part (a), it was pleasing to see many forming the correct distribution. Most managed to find $E(D)$ with only the occasional candidate making an arithmetical slip. Issues with finding the variance came from sign issues due to non-inclusion of brackets in the numerator or from missing out the squared on the numerator completely. Those who only achieved 2 marks, either did not state that the distribution was Normal - most often by omitting an N in front of their brackets, or they failed to divide their 3 by n for the distribution of $\bar{D}$. Some candidates tried to do all of their solution within their final stated distribution of the sample mean, whereas the most successful candidates broke the question down into finding the mean and variance separately for the continuous uniform distribution, followed by stating the distribution of the sample mean.

The most common score in part (b) was M0B1M1M0A0. Many failed to appreciate the answer from part (a) that $E(D)=x+2$ and as such when forming the equation for the lower bound of the confidence interval there was a mismatch in what values they were using. i.e. either 24.6 and 24.101 or 22.6 and 22.101 needed to be used. Ultimately many candidates ended up with a value of n of 3.19, which demonstrated similar good skills of manipulation, but unfortunately the same misconception. Some candidates failed to use the standard error in their calculation, although it was rare that the correct z value was not seen which was the source of one of the two marks typically scored. A few candidates did find 22.6 as the sample mean for x, but either ignored it and used 24.6, or made no further progress.