Examiners' Report
Principal Examiner Feedback

Summer 2025

Pearson Edexcel GCSE Statistics Higher Paper 2 (1ST0_2H)

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 2025
Publications Code 1ST0_2H_2506_ER

GCSE (9 – 1) Statistics – 1ST0 Principal Examiner Feedback – Higher Paper 2

Introduction

General comments

Candidates engaged well with the challenges presented in this paper, demonstrating a solid understanding of a broad range of topics from the specification. It was encouraging to see strong performance on questions involving standard techniques, such as interpreting comparative bar charts (Q03), analysing box plots (Q04), and calculating skew (Q10). Many candidates showed confidence in completing calculations and constructing diagrams, and demonstrated sound statistical reasoning when interpreting data.

As in previous sessions, certain topics proved more challenging, with candidates showing less familiarity with linear interpolation, comparative pie charts, and control charts. Centres should also remind candidates of the importance of clear, legible handwriting, as a small number of responses were difficult to read.

Most candidates were able to attempt the entire paper within the time allowed. However, some questions were left unanswered, and many responses were incomplete. Centres should emphasise the importance of presenting full working clearly, allowing examiners to award partial credit even if the final answer is not fully correct.

Question 1

(a) This question presented a challenge for some candidates as they began the paper. Several responses reflected varying familiarity with basic statistical terminology, with a selection of alternative terms such as ‘percentage’, ‘even’, and ‘binomial’ occasionally offered.

(b) This question was often answered without clearly explaining the difference between the two types of probabilities. Many responses included statements that did not meet the criteria for awarding marks. Explanations were frequently unclear or lacked detail, and some responses listed probabilities without any accompanying explanation. The most common mark awarded was for identifying that Peter’s probability was based on experimental data. Other responses included calculations of the difference between the probabilities or statements about which outcome was more likely; however, these were not credited unless they included reasoning to support the comparison.

Question 2

This question provided good differentiation and was well attempted by many candidates, with very few blank responses. Many candidates correctly identified that the hypothesis was written as a question rather than a statement. Most of the remaining marks were awarded for commenting on the reliability of collecting data from museums, the internet, or existing databases. Common uncredited responses focused on secondary sources being out-of-date.

A number of candidates suggested that primary and secondary sources should not be used simultaneously, which was not creditworthy. Two marks were awarded for recognising that a sample size of 2 per gender was too small, while noting that the overall sample of 100 was sufficient. Some candidates identified that splitting by gender was appropriate but rarely provided justification.

The question demanded structured, analytical responses. While the mark scheme was complex, it was generous to candidates who approached the task logically. The most successful responses followed a clear, step-by-step structure and addressed all aspects of the question. Many candidates, however, provided insufficient detail or poorly structured answers, often omitting key points. A frequent uncreditworthy response was the claim that sources were outdated.

Question 3

Parts (a), (b), (c) and (d) were answered well by the majority of candidates. In (b) incorrect answers neglected to write a comparison between the teams and instead wrote a comment on the trend for each team separately.

(e) Candidates were required to include the value read from the graph in their answer, or at least indicate clearly next to the diagram that the value had been correctly identified, by writing 52 (or 53). Some candidates assumed that simply stating "52" was sufficient, but it was necessary to write a correct comparison—namely, that 52 represented a smaller percentage of games won by the Lakers than the Clippers. A few candidates unnecessarily used these percentages to calculate and compare the actual number of games won out of 82. Marks were awarded where answers clearly demonstrated a correct reading of 52 (or 53) from the graph.

Question 4

(a) Many candidates correctly read a value of 23 or 24 from the graph, earning a method mark. Fewer then completed the method by subtracting their value from 80. A significant number misread the scale, recording 22 by incorrectly assuming each square represented 1. Additionally, some candidates subtracted their 23 or 24 from 90 instead of 80.

(b) Many candidates correctly identified that an advantage of using opportunity sampling is that it is quicker and easier to carry out compared to other methods. A common incorrect response, however, was to state that opportunity sampling is cheaper. While this may sometimes be the case, in the specific context of this question it was not necessarily true and therefore did not gain credit. Overall, most candidates demonstrated at least a basic understanding of the key advantage.

(c) Many candidates successfully created an appropriate diagram with the correct overall shape, and almost all plotted the minimum value and lower quartile correctly. Errors were most often seen in the placement of the median, upper quartile, or maximum, as these values were not directly aligned with a line on the axis. Difficulties in reading the scale accurately were also evident. A few candidates lost marks due to careless or ambiguous drawing of vertical lines.

(d) Candidates generally earned marks for identifying that the Yukon River had the higher median, while the Chena River had the higher range or IQR. Most correct interpretations were based on these measures. However, many candidates mistakenly believed that a greater range or IQR indicated that the fish were larger. Few candidates commented on skew, and when they did, it was often incorrectly identified or described using inappropriate terminology (e.g., "symmetrical skew"). Comparative language was another common issue: many candidates simply stated the values of the median or IQR, or assumed that using the word "whereas" constituted a valid comparison. Weaker responses frequently described the minimum and maximum values of the two datasets without evaluating or interpreting the range.

Question 5

(a) Many candidates earned the first mark for correctly placing 7 and 13 in the Venn diagram. Fewer scored the second mark for correctly calculating and placing 30, 10, and 9. The third mark was less frequently awarded, as many candidates did not write a number outside the circles to complete the diagram. The main difficulties arose when candidates had to calculate 30 and 9 and position them in the correct subsets. A large proportion failed to recognise that the totals in the subsets did not add up to 120, and did not correctly calculate the remaining 38 outside the sets but within the universal set. Some candidates did not score any marks, as they were unable to use the given information to find a correct starting value.

(b) Many candidates calculated P(G) correctly by determining the total number of musicians who play the guitar and dividing by the total surveyed. However, some relied on their completed Venn diagram from part (a) but had carried forward errors from that part. As follow-through was not awarded in the mark scheme for this question, only fully correct answers based on accurate values gained credit. Since the necessary information was clearly stated in the question, candidates were expected to use it directly and accurately rather than depending on incorrect earlier work. Follow through was awarded for the remaining parts of the question.

(c) Most candidates attempted this question, but as many gave 23/120 as an incorrect answer as those who answered correctly, often failing to work further with the fraction. Candidates who correctly identified the denominator as 40 generally went on to complete the question accurately.

(d) This question was generally not well answered. Marks were awarded for follow-through from candidates' Venn diagrams, but many responses consisted only of comments such as "both instruments cannot be used at the same time" or "playing one instrument does not prevent playing another." Responses that were purely descriptive, with no independence test calculated, received no credit. Given the structure of the question, marks could be awarded to candidates who noted that the events were not independent because part (b) did not equal part (c). Those who answered fully correctly typically applied the first method outlined in the mark scheme.

Question 6

(a) This item was answered correctly by the majority of candidates. Incorrect answers were usually around the size of the pie charts, suggesting that they may be different - rather than commenting on Marcus' conclusion based on the pie charts seen. Other incorrect answers were too vague with regard to the possible differences in the represented totals.

(b) Most candidates answered correctly, either showing full working or no working. Of those who did not secure both marks, the majority gave 468,356 for Norway instead of Finland. A small number incorrectly divided by 35 and then multiplied by 100.

(c) A number of candidates appeared to know the formula relating frequency and radius but were unable to apply it correctly in this question. Many did not take the square root or failed to recognise the squaring, often producing the incorrect answer of 31.8. Some candidates also inverted the two values in the fraction.

Question 7

(a) Many candidates scored full marks by correctly adding the five values. Those who did not typically made errors by performing unnecessary calculations after summing the values or by missing or including extra values.

(b) Linear interpolation has proven to be challenging for candidates. Many achieved the correct answer, but few showed sufficient working to secure the method mark in case of a calculation error. Most who did not score full marks missed part of the calculation, such as omitting the 30 or multiplying by 32.04 rather than by (50 – 32.04). Those who knew the correct method were generally successful, although some confused values, dividing by 17.76 or 1.56 instead of 19.52. Others made no correct start and attempted to find the mean instead. Overall, this question was widely attempted. Although blank responses were rarely seen, many candidates simply stated the group in which the median fell as their answer. Common errors included subtracting 13.80 from 45 instead of adding it to 30, producing an answer of 31.20. Despite these pitfalls, there were many good attempts on what is a challenging topic.

(c) Candidates needed to emphasise their comparison of the data in this question, and many responses were unable to be awarded full marks as they simply stated that they agreed with Kiran with no supporting evidence. Evidence of a comparison between the two medians was essential. Marks were allowed for follow-through for candidates who could not calculate the median using interpolation in part (b) but recognised that the 2023 median would fall in the 30–45 group, and therefore had to be greater than the 1960 median of 27. Follow-through was also awarded when candidates compared their part (b) median to support their reasoning.

(d) Simply stating that "linear interpolation is an estimate" was insufficient, as this had already been indicated in part (b). Candidates were expected to explain why interpolation is an estimate—either because the data is grouped or because it is assumed to be evenly distributed within the groups. Credit was given for comments noting that the website might use different-sized groups or may have access to the raw data.

Question 8

(a) Candidates correctly addressed the appropriateness of using a scatter diagram. However, many responses simply stated that it could be used to compare two data sets, which was insufficient for credit. Candidates needed to emphasise that the two sets of data are paired.

(b) Many responses incorrectly stated that height should be plotted on the x-axis because it is continuous data.

(c) This question was answered well by most candidates, with either "no correlation" or "positive correlation" accepted. While the strength of correlation is not assessed in the specification, it was encouraging to see many candidates describe the relationship as "very weak positive" or, as preferred, "no correlation".

(d) Most candidates correctly identified a negative correlation. Many also provided a correct interpretation, noting that as one variable increased, the other decreased—for example, as an athlete's long jump distance increased, their 100 m time decreased. However, some candidates confused their interpretation, suggesting that as the long jump decreased, race time also decreased. Others compared the PMCC from this part (-0.5) with that from part (c) (0.02), leading to confusing statements involving height, long jump distance, and race times. A

common error was generalising ‘worse performance’ instead of recognising that a lower race time indicates better performance. Additionally, some candidates thought negative correlation meant there was no link between the variables, or that as one increased the other decreased without considering the context.

(e) The expected response of “interpolation” was very rarely seen. Successful candidates were more likely to reason that the data for the given athlete was similar to that of another and use this to support their comparison. Many responses focused incorrectly on a supposed lack of correlation, misunderstanding the information provided. Others simply substituted the given value into the regression model to confirm the estimated result. Overall, the question was poorly answered. Many candidates relied on their (incorrect) answers from part (d) and did not demonstrate a clear understanding of what was required. A common approach was to substitute 8.39 m into the regression equation, obtaining 10.7 s—which was already given in the question—and then assert that the prediction was reliable. Few candidates referred to the table data to justify that both 8.39 m and 10.7 s fell within the observed values, which would have demonstrated interpolation. Very few mentioned sample size as a factor affecting reliability, and many made general comments about race times being similar without the specific reasoning required by the mark scheme.

Question 9

(a) Many candidates gave overly vague responses to this item. They needed to specify which control measure enhanced the reliability of the results, rather than offering a generic comment about extraneous variables.

(b) Most candidates scored either 0 or 2 marks on this question. Those who scored 0 either left it blank or attempted to use the formula for standard deviation. A common misconception was dividing by 2 instead of 3. The majority of candidates who scored full marks applied the first method outlined in the mark scheme. The charts on the next page showed slightly different values for the upper action and upper warning limits, but still had the same standard deviation between them. Candidates were awarded for the expected calculation, or if they used the information provided in the charts.

(c) Many candidates neglected to comment on the likelihood of this result, making parts of their responses ambiguous. While some correctly noted that action would be taken, others were premature in suggesting that the water supply would be stopped or the residents would have fallen ill.

(d) Most candidates answered this question well. Some candidates attempted to draw their own lower limits on the control charts and then concluded that values outside these limits required action. A few candidates commented that no values were “near” the warning limit rather than correctly stating they were within it, as the phrase “not near” could imply the values were either well below or well above the limit.

Question 10

(a) Most candidates correctly calculated the chain base index number of 99.1. Many then went on to calculate the geometric mean correctly, although a significant number added the chain base index numbers instead of multiplying them. A few made transcription errors or used only four numbers, while some incorrectly calculated the arithmetic mean. Of those who correctly

obtained the geometric mean of 105.5, most were able to state that this represented a 5.5% increase, although a few lost the mark for not specifying that it was per year.

(b) This question was generally answered well, with the majority of candidates able to calculate the skew correctly. However, some did not state whether the skew was positive or provide an accurate interpretation.

Question 11

(a) Candidates generally made good use of midpoints, but errors were common in incorporating the variable p correctly into the algebraic expression, with some writing forms such as "$2310 + 42.5 + p$" or other incorrect expressions. A correct denominator was frequently seen even in partially completed responses, although there was a wide spread of fully correct, partially correct, and incorrect answers.

A mixed set of responses was evident. The majority of candidates using a correct method were able to obtain the correct answer, though some mishandled the algebra and earned only the M1 mark for the numerator or M2 for a complete fraction. Candidates who scored no marks typically used an incorrect method for the mean, such as dividing by 5 or failing to incorporate p from the table.
Candidates who achieved full marks demonstrated a thorough grasp of the method and strong algebraic skills, successfully completing the calculation and earning the top score.

(b) Many candidates provided a correct and concise answer to this item. However, a significant number offered incorrect responses suggesting that the road safety officer was untrustworthy or that their equipment might be faulty.

Question 12

(a) Many candidates correctly calculated the IQR values for boys and girls, but were not always able to provide an accurate comparison. Some confused parts (a) and (b), offering an interpretation instead of the required calculation. Common errors included using incorrect values from the table or making numerical mistakes during subtraction.

(b) Candidates who scored well in part (a) often went on to score for the correct interpretation of their answer here. Follow through was also awarded for their answer in part (a).

Question 13

(a) This was a challenging question, but many candidates answered it well. The paper placed considerable emphasis on understanding the properties of the normal distribution and the spread of data within 1, 2, and 3 standard deviations. Many candidates demonstrated strong preparation in this topic and provided clear, concise answers.

(b) Responses varied, ranging from blank answers to attempts at a reasonably bell-shaped curve, with some candidates producing correctly drawn and labelled diagrams. Common errors included incorrect labels—such as using 1 SD, 2 SD, etc.—and completely inaccurate numerical scales. Of those who attempted a graph, the majority achieved the correct overall shape.

(c) Similar to Q10c, candidates often neglected to comment on the likelihood within their response. A good proportion of candidates were awarded this mark.

(d) Most candidates scored some marks on this question, with the histogram and comments about the mean and median for a normal distribution being the most commonly awarded. However, the section requiring comments on the tabulation was poorly answered—often skipped entirely or addressed with vague, irrelevant remarks. Many candidates also lost marks for the standard deviation, failing to quote specific values such as 68%.

Where marks were gained, they were typically for noting that the histogram allowed a visual assessment of the distribution or that the mean and median were approximately equal. Some candidates correctly cited standard deviation values to support their evaluation. A minority of candidates provided strong attempts, effectively commenting on both the tabulation and the plan, and achieved higher marks. The most successful responses were clearly structured, beginning with the table and then addressing the plan. These responses often commented that the class intervals did not overlap, were equally spaced, or that more intervals could have been included in the middle, while noting that the histogram showed the expected bell shape and that the use of one, two, and three standard deviations could illustrate the expected proportions of the data.

Common errors included focusing only on the plan and not the table, suggesting additional steps rather than evaluating the existing plan, failing to justify comments, or omitting the question entirely, possibly due to time constraints. Many candidates gained marks for discussing the histogram and two calculations, but few addressed the tabulation in sufficient depth. When class intervals were mentioned, comments tended to focus on their use for continuous data rather than evaluating the appropriateness of the distribution ranges.

Summary

Based on their performance on this paper, candidates should:

- Read each question fully and carefully before attempting to answer it. Check they have understood the requirement for comments in extended responses and try to give an answer for each section.
- It is not enough to simply read and state values from a graph or boxplot, without forming some sort of comparison and conclusion. Candidates should check the wording of the question too. Often, these values are not required and we are simply looking for visual comparisons to be made, simply stating “Yukon river had the larger median” was enough. Usually if a question requires these figures as part of your answer, the question will demand it e.g. “you should justify your answer”.
- Practise interpreting answers to statistical calculations and diagrams in context. E.g. when comparing the IQR, linking this to the data being less/more varied in the given context.
- Candidates should review interpretations of skew and revise the correct vocabulary when writing a comparison of skew for two different distributions.
- Candidates should practise identifying the appropriateness of calculations, diagrams and approaches in different settings
- Candidates must comment on the appropriateness of a conclusion if the question has asked you to. Candidates should not assume we know what their decision is with regards

to the appropriateness – we need to see a decision. ‘No/yes because’ is often the minimum required for these types of questions.

- Candidates should practise making comparisons, particularly those in context.
- Candidates should appreciate that an item requiring an answer worth two marks will require two parts to their answer. A comment with a conclusion/decision. Or a correlation type with an interpretation.
- Candidates should revise how to calculate geometric mean, and the median through interpolation.
- Candidates should revise how to calculate the radii for comparative pie charts.

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