



Examiners' Report Principal Examiner Feedback

Summer 2025

Pearson Edexcel GCSE Statistics
Foundation Paper 2 (1ST0_2F)

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_2F_2506_ER

All the material in this publication is copyright

© Pearson Education Ltd 2025

General comments

Most candidates handled the paper's challenges effectively and showed understanding across various areas of the specification. It was pleasing to see candidates performed well on questions requiring standard techniques such as tally charts, two-way tables, bar charts and probability.

Most candidates were generally confident with calculations and diagrams and showed good statistical understanding when interpreting them. However, they found questions that involved interpreting in context or evaluating methods a bit more challenging.

Candidates were generally able to complete the paper in the time given, though some left certain questions unanswered, including a few multiple-choice ones.

It's important for centres to remind candidates about the need for clear handwriting, as a few responses were difficult to read.

Question 1

In this first question candidates needed to complete a two-way table and then state a probability. This question was very accessible. Candidates were able to use the given cards to complete the two-way table in part (a) and then use the cards to state the probability in part (b). Most gave $\frac{4}{10}$ as an answer, with some going on to convert to a simplified fraction, or even a decimal or percentage, even though this was not required. A common mistake was from candidates writing $\frac{8}{10}$, which was obtained by adding the values in the table together and adding the total as well, therefore doubling the numerator, or by missing the white star and writing $\frac{3}{10}$ as the fraction. However, some did not read the question fully and gave the number of stars as an integer instead of as a probability.

Question 2

In this question candidates needed to discuss whether the visual representation was an appropriate way to represent the information that was given in a table. Most candidates scored at least one mark here for recognising that the visual representation was not an appropriate way to represent the information, and for giving an attempt at a reason. It is worth noting that to be awarded one mark they had to attempt a reason, and simply writing 'not appropriate' scored 0. Those candidates scoring the full two marks gave clear reasons often linked to the fact that there was no key or no years labelled. The word 'discuss' in the question may have led to some candidates trying to give a 'balanced' view rather than simply stating inappropriate / appropriate etc. saying why it is AND isn't appropriate. A small number of candidates misinterpreted the question and clearly thought it referred to the table. In these instances, they commented that it was appropriate as the years etc. were clearly shown.

Question 3

In part (a) of this question candidates were required to work out the mean number of films watched where the data was presented in a frequency table. Those that did have some understanding of the method would often, incorrectly, divide their summation by 4 (number of groups) instead of 12 (the number of students). It was common to see an incorrect method of dividing the sum of the frequencies (12) by the number of groups (4).

In part (b) they were given the mean number of films watched by 12 pensioners in the same week and asked to explain why Alina's claim may be incorrect. Candidates were often able to recognise that the data was only for films and not all television or discuss that they didn't know the length of the films or they may watch things on their computer, iPad, phone and not on television. Those that were less successful often gave answers that gave reference to the mean being a decimal, different sample sizes or reference to pensioners having more free time to watch films compared to students and not recognising that we did not have enough information about the total number of hours watched in that week.

Question 4

In this question candidates were given a histogram, pie chart and bar chart and required to state whether the mode, median or mean would be the most appropriate average to use to summarise the data. A straightforward question but, candidates found this question challenging and where a correct answer was seen it was often to recognise that only the mode could be found for the pie chart.

Question 5

Parts (a) and (b) of this question required candidates to use results from a spinner being spun by two different people to estimate probabilities. Candidates were more successful in part (b) of this question, where they were given the results of 50 spins in a table. In part (a), it was common for candidates to ignore the information given about Claire's spins, and for them to incorrectly write down the probability of spinning red ($\frac{1}{5}$) by looking at the spinner that was given at the start of the question.

Part (c) of the question scored well amongst the candidates, and they could identify that Dhiyan's results were more reliable to estimate probabilities as he had spun the spinner more times. Many chose Dhiyan but for other reasons such as 'he has put it in a table' but still scoring B1. A significant minority chose Claire, usually because it was easier than Dhiyan's table.

In the final part of this question, candidates were asked to decide, using Dhiyan's results, if the spinner was fair. Most candidates were able to state that the spinner was not fair but, were not always as successful in giving a reason why. Incorrect answers often seen stated the number of times the spinner landed on a colour, or that green was the highest and pink was the lowest and did not refer to there being significantly more or less of one colour or, that we would expect them all to be approximately equal of the spinner was fair.

Question 6

In part (a) of this question candidates needed to work out the median number of sheep for the data from 8 farms. Most candidates knew how to go about this and did so correctly. Careless ordering or simple arithmetic errors caused some to fail to get full marks but part credit for ordering, with one error meant they could score part marks. This question is an excellent example of where it is important for candidates to go back and checking their working.

In part (b) candidates needed to work out the new median if 2 more sheep needed to be included on every farm. Most candidates gave the correct answer here, though many recalculated all the numbers and went through the process again. Those who knew to merely add two to the median completed with less chance of error and, should they have made an error in part (a), could score a follow through mark.

In part (c) they were given the mean number of cows and pigs for the same 8 farms and asked to make a comparison. Most candidates were able to make a correct comparison and when marks were not awarded this was often for a response of 'the difference is 22'.

In the final part of this question candidates were asked which average was most appropriate to use with IQR. The median was the most common response but there were a good many who suggested the mean or mode, and some who offered 70 or 72 from parts (a) and (b) or gave the answer as an animal e.g. pig

Question 7

In this question candidates were presented with a grouped frequency table that showed the length of time, in days, that 15 butterflies had lived. In part (a) they needed to explain why the given mean of 4.5 days was an estimate. Many candidates were able to identify that the values used in the calculation were not the exact values but came from using the midpoint of each group. Where answered incorrectly, candidates often referred to the fact that there was an outlier or referred to the fact that there were three groups where the frequency was zero. Many candidates also referred the value given for the mean being a decimal and incorrectly stated that this meant that the mean was an estimate.

In part (b) they needed to explain why using the mean to represent the average length lived may not be appropriate. Candidates found this challenging, although a good number of candidates were still able to successfully identify that there was an outlier, or that the mean can be affected by outliers. Many candidates continued to refer to the fact that the mean was an estimate rather than why the mean was not suitable in this case. Many candidates also referred to the groups with a frequency of zero for this question.

Question 8

In this question candidates were presented with the scenario of a light bulb factory and needed to explain why conducting a random sample was more appropriate than conducting a census. Candidates found it difficult to connect the statistics to the problem and instead gave generic comparisons of a census and a random sample. It was rare to see a candidate reflect the problem in their answer and very few managed to realise that the items were single use and performing a census would destroy all the stock. When candidates were successful in gaining the mark, it was generally a generic comparison of a census and a sample, the most common successful answers were 'it takes less time' or 'costs less'. Unsuccessful candidates most simply stated advantages of a random sample, 'it's not biased' or 'everyone has an equal chance of being picked'. It was also common for the context to be completely ignored with answers like 'they might not answer'.

Question 9

In part (a) of this question, candidates were required to circle the type of data from a short list of possible answers. When an answer was selected, this was usually correct, although a significant minority gave an incorrect answer, predominately "quantitative". The tally chart and bar chart in part (b) was completed accurately by the vast majority of candidates. The scales on the graph seemed to help and there were relatively few incorrect answers such as drawing a bar with height 11 instead of 13.

In part (c) the majority of candidates were awarded a mark for the first bullet point, and "only tallied 55" was by far the most common response. Identifying that the animals were miscounted in some way was the second most common response. Other successful responses included comments of a more practical nature explaining why more animals had not been observed. Very few candidates used the words "biased" or "not random" in this practical setting. There were two marks available for giving two reasons and too many candidates gave only one. Centres should advise candidates, that in questions like this, they should aim to write two separate statements to maximise their chances of scoring both marks.

Question 10

In this question a teacher is investigating how many hour students spent revising for a test. In part (a) they needed to state one advantage of using a short questionnaire instead of interviewing each student. Candidates were generally successful and were able to give one advantage, commonly stating that it would be 'quicker'.

In part (b) they needed to state the type of question used in the questionnaire. The correct answer of 'open' was often seen but it was also common to see incorrect answers of other statistical words e.g. 'qualitative'.

Part (c) on cleaning data was answered correctly by some candidates with them commenting on the different units of time or inexact times. There were, however, lots of comments about putting the data in order or that it was confusing and not referring as to why the data needed to be cleaned.

Part (d) was a multiple-choice question where they needed to circle the type of diagram, candidates were often successful on this question and were able to correctly select 'scatter diagram'.

Question 11

In part (a) candidates needed to explain why the use of the pie chart may not be appropriate. It was generally answered well by candidates, many referred to not knowing the percentages of each department or that the sectors were similar sizes so difficult to decide which sector was the biggest. Of those who did not pick up the mark in this part, many did not make it clear why the pie chart was not appropriate and would make comments e.g. 'its unclear' or 'not accurate' or 'difficult to interpret'.

Part (b) was very well answered by candidates, and they were able to suggest a more appropriate diagram with the most common answer seen as 'bar chart'.

Question 12

In this question they were presented with a scatter diagram showing the marks that ten students received in Maths and Music tests. Part (a) was very well answered, and candidates were able to identify the outlier.

In part (b) they needed to describe the correlation shown in the diagram and most candidates were able to say that it was negative. When incorrect answers were seen it was often using words trying to describe a trend and not correlation e.g. 'decreasing'.

Part (c) was another successful part in this question. Many candidates plotted an appropriate line of best fit through the mean point graph. Where marks were lost it was due to not have their line going through the mean point or, lines were not splitting the data evenly. Many students again failed to use a ruler, and centres are advised to remind their students that the right equipment is needed to draw graphs accurately. The final part of this scatter diagram question was to suggest one adjustment that could be made to improve the reliability of any conclusions made using the diagram. Many candidates were able to correctly answer this by suggesting plotting more students on the scatter diagram, but it was also common to see candidates ignore the extra information given and suggest improvements to the scatter diagram e.g. 'draw another line of best fit' or 'remove the outlier'.

Question 13

In this question candidates were presented with a population pyramid for the number of managers in Australia from 2016 to 2107. In part (a) they needed to state the name of this type of diagram. Candidates found this very challenging, and it was common to see an incorrect answer that instead called it some type of bar chart. It was also common to see this question left blank.

Part (b) was a two-mark question and candidates needed to find the age group that the greatest proportion of managers are from. Many were able to score one mark by correctly identifying the age group of 45-54 and give a reason, often that it was because it was the largest percentage for males and second largest for female. It was not as

common to see candidates score 2 marks but justifying their answer with a correct percentage.

Question 14

This question was about a professor of sports science wanting to test if listening to music improved running performance. It was tested using a laboratory experiment. The context of this question helped candidates to be awarded marks when this is often seen as a challenge topic to foundation candidates.

In part (a) they needed to give two advantages of competing the study as a laboratory experiment. Candidates found this part the most challenging and often instead discussed the way the experiment was being carried out e.g. 'good to use males of similar age and fitness level' and not the advantages of it been completed in a laboratory. When correct answers were seen it was often that it would be less time consuming or that other factors can be controlled.

In part (b) they needed to state response variable and this was generally well answered, and candidates were able to state that it was the distance, or speed ran by the runners. When candidates were not successful it was often a response that included both distance/speed and music e.g. 'the distance they ran with music on'.

In part (c) they needed to explain how the professor could control the extraneous variable of listening to music or not. This part was answered very well with most candidates able to state that they needed to ensure all the runners listen to the same music, when incorrect answers were seen it was often for saying to allow the runners to choose their favourite music to motivate them.

Part (d) had mixed response. The most common correct answer was to make the test fair. Many responses recognised that extraneous variables impact results but did not gain the mark because they failed to comment that they must control them to reduce the impact. The most common mistake made was students commenting that it would make the results more accurate.

Question 15

In this question candidates were presented with a table showing the price index of eggs and sugar in the years 2010, 2015 and 2020, with 2020 been the base year for both ingredients. In part (a) they needed to calculate the percentage change for eggs between 2010 and 2020. Most candidates were able to interpret the index number to score one mark out of two for identifying a change of 29% but many failed to include the decrease or negative symbol.

Candidates found part (b) more challenging where they needed to calculate the price of 1 kilogram of sugar in 2020. Most common mistakes included using the egg value (71%) rather than sugar (77%), calculating 23%

The responses in part (c) were mixed, most either scored zero or two marks, as where they generally included a decision. Most correct responses mentioned inflation or that prices have increased since 2020 with several responses referring to increase since Covid 19 or referring to the cost-of-living crisis. The most common mistakes included referring to a continued decrease using information from the table or that prices varied in different shops/regions.

Question 16

This question was the first of the crossover questions with the higher tier and was an extended response question. It presented candidates with a plan for an investigation including Elises' hypothesis and a description of how Elise planned to collect data. The candidates were asked to discuss whether these plans were appropriate. It was pleasing to note that most candidates attempted this extended response question, although there were still some blank responses.

Many candidates were able to gain some marks, but the majority did not realise that five marks required five statements and, in some cases, did not recognise that they should comment on both the hypotheses and the collection of data.

Common correct observations from candidates for the hypotheses plan recognised that it was not suitable as it was a question and should be a statement, although many candidates did think that it was a suitable hypothesis.

For the collection of data, correct observations generally included that using secondary data or collected data from a museum was not appropriate as they may be unreliable/inaccurate or that it would be quicker/cheaper to use this secondary data. Many also recognised that collecting a sample size of 100 was appropriate since it was a large sample size or that only collecting data for only two males/females was not appropriate as it was a small sample size. They also made comments about Elise not collecting data on body mass as per her hypothesis or that she should also collect data for young mammals as well as adults. Many candidates referred to collecting secondary and primary data and that mixing them both was not appropriate which did not receive any marks. It was rare to see candidates identify that the data may need to be cleaned, or they may be missing data or that there was no indication of how the sample of skulls from the museum would be collected. Many candidates made comments about the secondary data or museum data been out of data which was not awarded any marks in context of this question.

Question 17

This was the second of the crossover questions with the higher tier and foundation candidates were again successful in been awarded marks throughout the question. In part (a) they needed to complete a sentence by adding an appropriate statistical word. Most candidates could correctly write 'secondary' and be awarded this mark.

In part (b) candidates were presented with a dual bar chart for the number of games won by the Clippers and the Lakers basketball teams for four seasons from 2015 to 2019. Candidates were very successful making one comparison between the two teams and gaining this second mark, often with a response that identified that Clippers won more games. When incorrect answers where seen it was often candidates comparing just one season or for comparing each team's performance from one season to another.

In part(c) they needed to describe the trend in the number of games won by the Lakers over the four seasons. Again, candidates were very successful in this question and could correctly identify that they trend was increasing or that the Lakers were winning

more each season. It was rare to see an incorrect response that included the word 'correlation' like in previous years.

Part (d) was a two-mark question where candidates had to complete a table showing the cumulative percentage of games won by the Clippers. Candidates were not as successful in this part of the question. It was also common to see candidates get one of the figures correct and the other incorrect even though both figures involved the same calculation.

In the final part of this question there was a line graph showing the cumulative games played and the cumulative games won by the Lakers in the same season as part (d). Candidates needed to decide which team won the most games by the end of the season and justify their answer. Most candidates could correctly identify that the Clippers won the most games but were not able to justify their answer with figures which meant that they were not awarded this mark.

Question 18

This was the final common question with the higher tier. In part (a) candidates had to use a cumulative frequency diagram to make an estimation. They were often able to read off a correct value for the number of pikes that had a length of 440 mm or less but would then give this as their final answer instead of using it to estimate the number of pike there were 440 mm or more i.e. subtracting it from 80. Of those that did realise they had to do a subtraction they occasionally subtracted incorrectly from 90, the largest value on the y axis and not 80, the number of pike caught. Some candidates struggled to read 440 mm off the scale from the graph axis.

In part (b) they needed to state an advantage of opportunity sample in relation to the collecting of the data for 80 pike. Many candidates were able to give one of the advantages stated on the mark scheme with 'quicker' or 'easier' as the most common questions. It was also that some foundation candidates did not understand what opportunity sampling was in the context of sampling fish and would give advantages that related to sampling people e.g. it's cheaper or just give a standard answer of 'it's not biased' which were not awarded the marks.

Part (c) was answered well, and it seems that candidates have been well prepared in drawing box plots which was pleasing to see. Many interpreted the given information and plotted these correctly to score full marks. A few candidates made plotting errors and when this occurred it was due to reading the scale incorrectly especially for the median and maximum point.

In Part (d) candidates needed to compare the two box plots and interpret one comparison in context. It was pleasing again to see that most candidates were able to attempt this question, and blank responses were not as common as in previous years. Although when candidates did attempt this question, their response was often not awarded any marks due to them comparing the maximum/minimum values and the lower/upper quartiles instead of comparing the medians and IQR/range and skew. Generally, candidates were able to compare the medians, but a few just quoted figures without using comparative language. Where candidates had correctly compared the median, they then struggled to interpret this in context or would state that the 'Yukon River is longer', which was not awarded the mark as it was an interpretation about the length of the river and not the pikes. It was pleasing to see, unlike previous years, the

use of the word 'medium' for median or incorrectly calling the median the 'average' or the 'mean' was rarely seen. Comparison of the range/IQR were less successful than those that compared the median, but like the median some just quoted the figures. Very few candidates were able to identify the different types of skews displayed in the two box plots.

Question 19

The final question on the paper was a triple Venn diagram and generally candidates were able to be awarded at least mark.

Part (a) was the most successful where candidates had to calculate the probability that the musicians play all three instruments.

In part (b) they needed to calculate two missing values in the Venn diagram. Many candidates were successful in finding at least one of the figures, mainly the value of x but then went on to incorrectly find the value of y as '45' and forgetting to subtract the '30' from their calculation for the students who did not play any of the instruments.

In part (c) candidates were asked to calculate conditional probability. This proved challenging for most candidates. Many did not understand the notation and instead gave a figure off the Venn diagram and not a probability or failed to recognise that it was conditional probability. Incorrect answers often seen were using the figure from the Venn diagram for a musician playing both the guitar and piano e.g. '5' or ' $\frac{5}{120}$ '

Summary

Based on their performance on this paper, candidates should:

- Ensure the correct statistical language is used throughout when making comparisons, stating figures, or using the word 'whereas' is not a comparison.
- Know to compare the median, IQR/range and skew when presented with box plots.
- Practice interpreting in the context of the question.
- Practice deciding when it is appropriate to use which average for different sets of data.
- When presented with a scenario and asked if the data collection is appropriate, ensure that they consider the scenario and note any advantage/disadvantage of the method given.
- For extended response questions, check the number of marks that are available and ensure that they make a corresponding number of points in their response.
- When deciding if a data collection is appropriate, ensure that they give a reason why – merely stating it is not appropriate is not enough.
- If a question asks if something is appropriate, they need to ensure that they state their decision.
- Candidates should revise conditional probability, and the notation used.

