



Pearson
Edexcel

Examiners' Report
Principal Examiner Feedback

Summer 2024

Pearson Edexcel GCSE
In Statistics (1ST0)
Foundation Paper 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 2024

Publications Code 1ST0_2F_2406_ER

All the material in this publication is copyright

© Pearson Education Ltd 2024

GCSE (9 – 1) Statistics – 1ST0

Principal Examiner Feedback – Foundation Paper 1

Introduction

General comments

Candidates responded well to the challenges within this paper, demonstrating understanding of a wide range of topics from the specification.

It was pleasing to see students performed well on questions requiring standard techniques such as tally charts (Q03), comparing two means (Q04) and reading a choropleth map (Q10). They were generally confident at completing calculations and diagrams and demonstrated good statistical understanding when asked to interpret these. Like previous exam sessions candidates were still less familiar with certain topics e.g. stratified sampling, relative risk and cumulative frequency curves. It is worth mentioning here that centres must impress upon their students the importance of legible handwriting as a few responses were very difficult to read. Also, the use of phrase ‘more accurate’ is too often seen, and not usually awarded any marks. Students were generally able to attempt the whole paper within the time allowed, although some students made no attempt on some questions and many responses were left blank. Centres should also encourage students to show full working and set this out clearly so that partial credit can be awarded if a fully correct solution is not obtained.

Question 1

This question was on probability. This question as a whole was very accessible. Part (a) of the question was multiple choice and required candidates to identify an even outcome from a set of 10 cards. Many candidates selected the incorrect answer of ‘likely’.

In part (b) candidates were required to select the two outcomes that were equally likely. This part of the question scored well, better than part (a).

Part (c) and part (d) required candidates to indicate on a probability scale, the outcome of two different events. Part (c) was for a single event, and part (d) was for a union of events. As expected, part (c) scored better than part (d).

Question 2

This question required candidates to read information from a simple bar chart. The bar chart displayed information on the number of days a fitness group, with a population of 40, attend a gym each week. In part (a) the candidates were required to find the probability that a randomly chosen person attends the gym exactly 2 days per week. The majority of candidates scored this mark. A common mistake made was the frequency being given instead of the probability of this outcome.

In part (b) of this question, candidates were required to state the modal number of days of attendance to the gym. This item scored very well amongst the candidates.

Part (c), candidates were required to critique a conclusion and explain why this might be incorrect. A good number of candidates were able to correctly articulate that the majority of the fitness group attend 2 or more days a week as an explanation as to why the conclusion was incorrect. The quality of written responses did vary though. There was ample flexibility in the mark scheme on how this could be written, with clear intention of a correct answer being awarded. We condoned responses referencing a small sample size, despite the bar chart representing a population of a fitness group. A common mistake seen was candidates occasionally wrote that ‘Jenny’ was incorrect because eight people attended on three days. We did not allow comments suggesting that the usage varied week by week. Some candidates did not use the data from the graph to counter the conclusion that had been drawn, giving throw away comments such as ‘increase the sample size’ with no obvious link to ‘Jenny’s’ conclusion or the bar chart.

Question 3

This in part (a) of this question, candidates were required to complete a simple tally chart on the number of medals won by British swimming Olympians. This part of the questions scored very well, with the majority of candidates scoring two marks.

In part (b) candidates were asked to suggest a suitable diagram to display the results of the tally chart. Many candidates stated the expected answer of ‘bar chart’ but there were many other correct alternatives accepted. This part of the question scored very well.

Parts (c), (d) and (e) were marked together. In (c), the candidates were asked to state the mode (or modes) of the number of medals won by British Olympic swimmers, and this scored well both in accuracy and in follow through from the candidates tally chart. A common mistake seen included the number 2 being written instead of the correct answer (1 and 3), because the candidate had counted the number of modes, or because the frequencies in the table happened to have three 2's. Another common mistake was a mode stated as 6, because the two modes (1 and 3) had this as their frequency. Part (d), candidates were asked to work out the median. For this part, no follow through was awarded from their table as the candidates had access to the raw data. This part of the question did not score as well as expected. The mistake most commonly seen was candidates using their frequencies instead of the raw data to find the middle number. Part (e), the candidates were asked which average, mode or median would be the best to use for summary. Given the context, the candidates were expected to state the median would be best because there were two modes. Follow through was awarded in this question dependent upon how many modes the candidate stated in part (c). There was lots of flexibility on the award of marks in this part of the question, enabling many students to still gain marks even if an error had been made in stating the modes, or in their calculation for the median.

Question 4

Part (a) of this question was to find the probability of an event happening from a small list of possible outcomes. This part of the question proved to be accessible to the majority of candidates, however, a disconcerting number of candidates made no attempt. This was perhaps due to the less-familiar requirement to find a probability from a set of figures like this.

In part (b), candidates were required to assess the validity of a conclusion made about the average heights of two different sets of pyramids. The first two marks were awarded for calculating the mean of a small list of data. Many candidates scored these two marks or higher. The third mark for applying an appropriate adjustment to the calculated mean, or to the average of the second set of data in order to make a comparison. The final mark was for a correct conclusion based on this clear comparison. Some students receiving only two marks neglected to make their comparison figures clear for their conclusion. To compare the two sets of pyramids a lot of sensible approaches were taken, including a minority who compared the scale factors involved to make their conclusion. The majority of candidates tackled this question as the mark scheme anticipated, doubling the given mean of the second set to compare with their calculated mean from the data given.

In part (c), candidates were required to work out the largest value in a set of data when the lowest value and the range is known. A significant number of responses demonstrated a lack of confidence with this alternative approach to a range calculation, but overall a larger majority of candidates did answer this item correctly.

Question 5

Question 5 was an extended response question. Candidates were required to critique a plan to investigate the length of time a learner waits for a driving test. The candidates were prompted to comment on three areas: data collection, calculations and diagrams. For data collection, many scoring candidates commented on the inappropriateness of collecting data solely in June because this would not be representative. The mark scheme also allowed for the converse of

this, that collecting data solely in June would be appropriate for consistency – although this was not as common an answer amongst candidates. Candidates struggled more with commenting on the calculations proposed. We were looking for a comment on the average being appropriate as a summary statistic for each of the regions, or that the plan does not specify which average Claire is planning to use so inappropriate. Some candidates are confusing ‘average’ for mean, so did not notice that the type of average had not been specified. This was often given away by candidates writing the average includes all of the data (including outliers) so not appropriate. Many more candidates did manage to comment on the appropriateness of using the range as a way of judging the spread of data for each region. For diagrams, we were looking for comments along the lines of the bar chart of the average waiting time is appropriate because comparisons can be made. Or conversely, the bar chart is not appropriate because the data is continuous. Comments on the pie chart being inappropriate because ranges are not easy to compare in this way were seen, but occasionally the candidate was a little too vague in their meaning. It was unfortunate to see so many blank responses. This extended response question was designed with foundation candidates in mind and thought to be very accessible prior to sitting. It was a shame to see candidates not scoring more on this question. It's important to note the candidates must conclude on their comment(s) as to whether the method was appropriate or not. Some candidates wrote very long answers discussing the data collection being inappropriate but made no comment on the calculations or diagrams as prompted by the question design.

Question 6

Part (a) of this question was to complete a composite bar chart. This proved very accessible. We allowed for non-ruled bars. Both heights needed to be correct for any marks to be awarded. The second mark came from correct shading. We allowed for messy shading in the bottom section of the bar, the diagonals in the top section had to be in the correct direction. There were some non-attempts at this item, which may be due to this part of the question being overlooked by the candidates. In Part (b) candidates needed to state that the overall height of each bar represented the total number of points. Some candidates stated more detailed answers such as the number of points scored in this season. The minimum requirement accepted was ‘total points’. In (c), many candidates scored the first mark for comparing the overall points scored by the two teams. Others, tended to score 2 marks for comparing the points from draws and the points from wins correctly, but forgetting to compare the overall total score. Too many candidates just stated the number of points for draws, wins and/or overall and forgot to actually compare them so scored no marks. Please can centres remind their candidates that it is not enough to simply write 33 and 35 without concluding that one team scored higher with 35. This question did not require the actual numbers to be given in the answer, we needed to see conclusions from comparisons made.

Question 7

Part (a) required candidates to state what the population was for this investigation. Many candidates wrote ‘the visitors’, but as emphasised in previous mark schemes and examiner reports - we needed to see ‘all the visitors’. The emphasis on everyone that visits the theme park was required. If the candidates had been asked what the sampling units were then ‘the visitors’ would have been appropriate. Parts (b) and (c) centred around convenience sampling and giving a disadvantage to using this method. It proved difficult for some candidates to

articulate what a convenience sample is. We allowed answers in context the question such as 'he asks only people queuing up for food at the time'. We were looking for 'easy to access' sample, not it is 'easier to conduct' a convenience sample. In part (c), we did not allow 'not fair', we were looking for not representative/biased. Part (d) required students to comment on a data collection table. Please note, in this series we allowed for the 'table should include more options', and 'the option 'other' should be included' - as two separate marks. This is not usual practise, and would usually only scored one mark for either one of these comments being made. Many candidates incorrectly commented on the table not having enough room for tallies, that the table was missing its frequency column, or the tallies would be too difficult to count up at the end of the sample. In part, the candidate needs to conclude that the use of a stem leaf diagram was not appropriate and make an attempt at a reason for the first mark. For both marks, we needed that this was not suitable because the data is qualitative (not numerical/quantitative). In parts (d) and (e) a large number of candidates scored the first mark, and a small minority scored two marks.

Question 8

Question 8 was centred on a scatter diagram, and information with regards to the weights and heights of rugby players. In (a) the candidates needed to give a reason why a scatter diagram is appropriate, and we were looking for 'the data is bivariate' or 'because scatter diagrams show correlation/relationship between two variables'. The mark scheme details the wording required. Many candidates incorrectly stated that the scatter diagram was appropriate because the data is continuous and many more left this item blank. (b) – (d) were marked together. In part (b) the candidates were asked to plot 3 points into the scatter diagram. The majority of candidates managed to score one or two out of the possible two marks here. There were still a large number of candidates who made no attempt at this part of the question or read the scales very poorly. Those who made no-attempt, it is possible they did not read the question carefully enough and this part went unnoticed. Centres should ensure candidates are told to check and read every part of the question. Part (c) asked candidates to name and interpret the correlation seen in the scatter diagram. Large number of candidates were able to identify the positive correlation, and many went on to correct interpret this with regards to the context of the question. Some scoring just 1 mark forgot or were unable to interpret the positive correlation. In part (d) the candidates were required to draw a line of best fit through the double mean point. Many candidates scored one mark on this item for either a good line of best fit but missing the double mean, or the double mean being plotted and the line of best fit missing or poorly drawn. Candidates must ensure they use a ruler to draw their line of best fit. Part (e) proved tricky for candidates. They were required to give a contextual interpretation of a Spearman's rank correlation coefficient with a value of 0. Too many candidates wrote 'no correlation'. In part (f) and (g) candidates were given a table with the Spearman's rank correlation coefficient and the gradient of the lines of best fit for two sets of data. In (f), the candidates were to compare the two correlation coefficients and conclude that one was bigger than the other, so the heights and weights for that team were more closely related. Many candidates scored the first mark by saying that the England team had a higher correlation coefficient than Wales but failed to give the context. Others confused the question and seemingly compared each teams Spearman's rank correlation coefficient the same teams' gradient of line of best fit. So comparing across the table instead of down. In (g), we were looking for assessment of the validity of the conclusion with supporting reason for two marks, or a comparison of the gradients with no or incorrect assessment of the validity. Too many candidates used a poor descriptor for the comparison of the two gradients.

An answer of England had a higher gradient than Wales was not correct. We were looking for 'greater/steeper'.

Question 9

Part (a) on cleaning data was answered mostly correct with candidates commenting on the different units of time and missing data. There were, however, lots of comments about why you clean data that were not specific to the data provided e.g. 'to remove errors' 'to make it accurate', 'to make it easier to process'.

Part (b) required students to comment on the appropriateness of using just one box plot to compare two sets of data. Many candidates did not attempt this question. Frequent errors included not accurate/ has decimals so cannot plot/ hard to understand/ stating "no" but not giving reason. Some were able to identify that data couldn't go into a single box plot but could not explain why. There were also some candidates who tried to suggest an alternative diagram that could be used. Centres should remind candidates that offering an alternative rather than commenting on the method suggested will not usually earn them marks.

Question 10

In part (a), of this question candidates were to comment on the appropriateness of using secondary data. It was generally answered well by candidates. Many made reference to it being a sensitive question, however most scored B1 for stating that it would be 'easier to collect' or 'faster' to collect secondary data. There were also converse statements about why primary data is not appropriate. Of those who did not pick up any marks, many did not make it clear whether they were talking about primary or secondary data and some had vague and incorrect comments about secondary data being 'more clear' or 'more accurate'. Part (b) was answered very well, with a small number of candidates writing 'Virginia' instead of 'West Virginia' for one of the three states required.

In part (c), candidates were required to comment on the validity of a conclusion with regards to the location of the highest-income states. Some candidates left the question blank or simply stated 'yes' or 'no', many attempted the question and were able to pick up at least B1. The candidates who did not score any marks seemed to repeat what was in the question and did not make enough of a reference to the choropleth map. However, most of the responses were worthy of at least B1. For those who scored B1, many did not assess the conclusion, or some picked up the SCB1 mark for comparing the East and West, as they did not fully understand what the question was asking of them. For those who scored B2, there was a mention of the darkest shading or over '100 thousands' with a correct conclusion. Some mentioned individual states on each coast, and some others claimed that the conclusion was not correct due to Colorado being a high-income state. Centres should remind candidates to check that their written response is not simply a rewrite of the question, especially in validity questions such as this. Evidence must be given or described in their response to guarantee the marks can be awarded.

Question 11

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. More often than not, the question was not attempted. Few circled an incorrect answer if attempted.

Parts (b) and (c) were marked together. In (b), candidates were required to draw a boxplot from a table of summary statistics. Some adjustment was needed to be made to this information in order to plot the box plot. This was missed by some candidates, as was evident by the attempts seen whilst marking. Provided a box plot with two whiskers was drawn the 1st B1 mark was often awarded for a correct largest value drawn for the upper whisker. B2 required the upper quartile or the lowest value to be plotted correctly and was given for a line drawn between two grid lines. B3 was awarded to many candidates, but awarding for accuracy of the boxplots was difficult for those candidates using a blunt pencil and drawing thick lines or not using a ruler.

Part (c) was an extended response question and required the candidates to compare the two/three box plots from part (b). Many students did not attempt this question and left it blank. When students did attempt this question, their responses were often not awarded any marks due to them comparing the maximum/minimum values and lower/upper quartiles instead of comparing the medians and IQR or range. The question did not require context, just interpretation. This made the mark scheme more generous than usual with regards to the comparison of the medians. When centres come to review this question, please bear in mind that the comment ‘international had the highest hand span’ was awarded with benefit of doubt, positively marking, due to the assumption the candidate meant that this was because the median value for international was larger, not because it had the largest hand span. The highest mark awarded on this item was usually B4, with very few candidates able to identify the different types of skew displayed in the two/three box plots. The mark scheme allowed for two different comparisons of spread, IQR and range – but only on one mark for an interpretation of spread. It is worth mentioning again here, that some candidates are not comparing the data presented – they are just stating values for the median/IQR etc. They must make a comparison of the data they write in their responses. Some candidates are using medium instead of median, or incorrectly calling the median the ‘average’ or the ‘mean’.

Part (d) was on stratified sampling. In this series it was worth three marks, but in previous series it has been linked to extended response questions. As seen in previous series, stratified sampling is not understood by the majority of foundation candidates. Many candidates did not attempt this item at all. Those who gained B1 or B2 were able to state the calculation required and an accurate sample size for one of the groups (for B2). Very few managed to gain all three marks for describing the random selection process from these groups once the sizes have been determined.

Question 12

Question 12(a)(b) was on relative risk, calculating and then interpreting. While a good number of responses showed they could express the risk in an appropriate way, the large majority were not able to calculate the relative risk correctly. Many achieved just one in four marks for this question for calculating one probability correctly. Very few went on to compare the size of one probability to the other. A small number of candidates gained two marks for the use of ratios. In part (b) candidates were only awarded a mark for a correct interpretation provided they had

achieved a distinguishable number in part (a) that they had tried to use for the interpretation. Those who had a figure could then interpret it fairly reliably but too often were commenting on the number of fails or passes rather than the risk of failing.

Question 13

Question 13 was centred around a cumulative frequency curve. In part (a), candidates were required to read the median value from the graph. When attempted candidates managed to state a correct answer within the expected range. This mark was not awarded for a value given within the range, but incorrect working seen on the graph. Occasionally, candidates stated an answer that was too large (52) due to halving 800 rather than 650.

In part (b) most candidates were able to read the correct values required from the graph and calculate the difference, achieving 2 marks. Errors were made calculating the percentage using the wrong divisor (800 the most common) and quite a lot of comments at the end referred to the number of members rather than the percentage/proportion that were between the ages of 45 and 60 in each parliament.

Summary

Based on their performance on this paper, students 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.
- Check graphs carefully for any part of the question asking you to complete it. Scatter diagrams will often (but not always) require additional points to be plotted. A bar chart with a blank space in the middle or at the end is likely to mean there is a requirement for them to complete it. Please encourage candidates to check where all marks are being awarded in a question.
- 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. Usually if a question requires these figures as part of your answer, the question will demand it.
- Candidates should practice interpreting answers to statistical calculations and diagrams in context. E.g. They should upgrade their responses to always comment on the comparison of IQR and the link to the data being less/more varied with context.
- Candidates should practice identifying the appropriateness of calculations, diagrams, and approaches in different settings.
- Candidates must comment on the appropriateness of a conclusion if the question has 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 practice on 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 and interpret relative risk.
- Candidates should revise how to calculate the proportionate sizes for stratified sampling, and practice describing how to conduct this type of sampling method randomly.
- Candidates should revise cumulative frequency curves and reading values from it.

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