



Pearson
Edexcel

Examiners' Report
Principal Examiner Feedback

Summer 2024

Pearson Edexcel GCSE (9-1)
In Statistics (1ST0)
Higher Paper 2

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_2H_2406_ER*

All the material in this publication is copyright

© Pearson Education Ltd 2024

General comments

Most candidates have responded well to the challenges within this paper and demonstrated an understanding of a range of areas of the specification. It was pleasing to see candidates performed well on questions requiring standard techniques such as cumulative frequency diagrams (Q04), Spearman's rank correlation coefficient (Q06) and Peterson's capture re-capture (Q10).

They were generally confident at completing calculations and diagrams, also demonstrated good statistical understanding when asked to interpret these. Candidates found questions requiring interpretation in context and evaluation of approaches or techniques slightly more challenging.

It is pleasing to report that questions requesting candidates to explain, assess or interpret have improved, compared to previous years.

It is worth mentioning here that centres must impress upon their candidates the importance of legible handwriting as a few responses were found very difficult to read. Candidates were generally able to attempt the whole paper within the time allowed, although some candidates made no attempt on some questions.

Question 1

In part (a) of this question, candidates answered well. They were able to state why using secondary data was more appropriate or referring to the disadvantage of trying to collect primary data. Some candidates lacked the clarity on whether they were referencing secondary or primary data in their answer which is where they lost this mark. There were a small number of candidates who scored no marks for indicating that secondary data is more accurate.

In part (b), the vast majority were able to identify the three states having the lowest mean income. There were few responses that did not achieve the mark.

Candidates answered part (c) well, with some referring to the darker shading or more areas with ≥ 100 on the East and West coasts. Some answers lacked precision when referencing the specific features of the choropleth map, often in such cases, just restating the question rather than referencing the shading on the map or interpreting the key. Some candidates misinterpreted the question and thought that they had to compare the East and West coasts, and concluded the East coast has the highest income states. Once again it is important to stress that where a decision is required, this must be provided as well as the supporting evidence for the decision.

Question 2

In part (a) candidates were asked to decide which of the statistical words given best described hand span as a type of data. As would be expected, this was answered well by candidates.

Part (b) was answered well, and it seems that candidates have been well prepared in drawing box plots. Many interpreted the given information to find all 5 values and plot these correctly to score full marks. A few candidates made plotting errors and when this occurred it was due to usually reading the scale incorrectly.

Part (c) saw a varied range of marks but seemed accessible to all candidates. Generally, candidates were able to compare medians, but a few just quoted figures without using comparative language. A few showed to be confused as to what average the box plot used and so quoted average/mean. Again, many were able to compare range and/or IQR but, like the median, some just quoted the figures. Those students that compared skew either did so correctly or made the usual errors, which included confusing positive and negative skew. Candidates that compared the largest/smallest and were not given credit for these comparisons. When an interpretation was given, the most common correct answer came from interpreting the medians with many candidates saying that the international level had the largest hand spans. Those that interpreted the range/IQR were less successful (IQR was done better than range) than those that

interpreted the median. A correct interpretation of the skew was rarely seen.

In part (d), candidates needed to explain how Pavel can obtain his stratified sample. The majority of them were able to show or describe how to calculate the number to be sampled from each stratum and give at least one correct figure. Candidates were less successful in then explaining how they would then randomly choose these students from each stratum. Those that didn't understand what a stratified sample was, would describe a random sample or suggesting taking $20 \div 5 = 4$ from each stratum.

Question 3

This question related to the calculation and interpretation of relative risk. This proved challenging for the majority of candidates.

Part (a) required candidates to calculate the relative risk of failing the skills test having taken course A compared to course B. This question was answered correctly by some candidates, while others subtracted probabilities instead of dividing them getting an incorrect answer of 0.1

Part (b) of this question required candidates to interpret the relative risk that they had found in part (a). This was not answered well by many candidates. For those candidates that got the correct answer of 0.75, in part (a), they then misinterpreted this answer and said this was a 75% chance of failing the skills test having taken test A. Many instead stated that test B was more difficult and showed no understanding of what relative risk was.

Question 4

Part (a) of this question required candidates to complete a cumulative frequency diagram using the information in the table. Generally, the question was well answered with most candidates correctly plotting the upper boundaries and joining to make a curve. Although some candidates lost a mark by having a positive and then a negative gradient between the last two points. A minority of candidates failed to plot the upper boundary and plotted the points at the midpoint of the intervals given in the table.

To answer part (b) of the question candidates needed to read appropriate values from their cumulative frequency graph and find the difference between the two. This part of the question was answered well, even by those candidates who made mistakes in part (a). A small number of candidates read the scale on the cumulative frequency diagram incorrectly.

Question 5

Part (a) of the question required candidates to complete a back-to-back stem and leaf diagram. The majority of candidates were able to construct an ordered stem and leaf correctly, it was rare to see the numbers unordered

or numbers missing. Completing the key for their diagram proved to be challenging with many candidates only displaying a key for the right-hand side of the diagram or attempting to show the 2 different keys but failing to indicate how these related to the two sides of the stem and leaf.

In part (b), candidates were asked to calculate the median. A common incorrect answer was 96 from counting outwards towards the stem on the second line of the Welsh side of the stem and leaf diagram instead of from the stem outwards to get the correct answer of 92. Occasionally, candidates incorrectly calculated the England median.

In part (c), candidates were asked to calculate the interquartile range. Candidates usually calculated the IQR correctly using 98-84, although many candidates could only identify the lower quartile correctly.

Part (d) of the question presented the candidates with summary statistics and conclusions about the weights of the England Rugby Union Backs and asked if the statistics supported the conclusions. Candidates were generally able to compare at least one of the medians or the interquartile ranges, and often correctly compared both. However, many candidates would often just state the values of the statistics and failed to make a comparison. A few candidates also confused themselves with which showed consistency, the median or the IQR or would use the incorrect statistical words e.g. mean instead of median.

Part (e) asked candidates to comment on the appropriateness of using the median and interquartile range in an article. Many candidates could state that it wasn't appropriate because 'people won't understand the IQR' or that median and IQR are not affected by outliers. When incorrect answers were seen it was candidates instead describing what the median and IQR showed.

Part (f) asked candidates to give a limitation of using the data provided to compare all the players on the two teams. Many candidates failed to explain this correctly, either resorting to giving disadvantages of using the median and/or interquartile range. Some were able to refer correctly to only been given data for the backs or one position of the team. Information about the data, for some, researching and investigating the effect of changing land usage on the amount.

Question 6

Part (a) presented data on surface runoff water for areas of land. Candidates were given the ranked data for change in forest cover area and were asked to calculate Spearman's rank correlation coefficient for the information in the table and provide an interpretation of the correlation in context. The majority of candidates could make a start to this question by completing the ranks for change in surface area runoff and then finding $\sum d^2$.

Many candidates made mistakes when ranking the 15 pieces of data but were still able to find $\sum d^2$ for their ranks. Where an attempt to add the

values of d^2 was made, it was pleasing to see this was often followed by use of the correct formula in an attempt to find Spearman's rank correlation coefficient. Candidates generally used the formula given with a high degree of success, but some forgot to subtract from 1, squared the individual ranks and some had an incorrect value for n .

The candidates were more successful in giving an interpretation of the correlation coefficient by giving the type of correlation. Although, a few candidates mixed up their interpretation and instead would give the answer as positive or negative skew. Candidates were less successful in giving an interpreting of their correlation coefficient in context and this was often seen as 'change in forest cover and surface run off disagree' instead of the correct interpretation of 'as forest cover area reduces the amount of surface runoff increases'.

Question 7

The interpretation of the index number was well understood by the vast majority of candidates in part (a) with many scoring both marks by identifying both the increase and the 8.9%. Those who scored only one mark often recognised the increase, but incorrectly interpreted the value of 108.9 – common error was 108.9%. A small minority of candidates did not make reference to the 108.9 at all, just stating that the CPI had increased. Others incorrectly compared 2019 and 2020 and not recognising that 2015 was the base year and then gave 1.1% as the percentage increase. Some candidates tried to reason why there was an increase (which wasn't required), usually indicating that 2020 was the pandemic.

Part (b) was generally well answered, candidates were able to find the percentage increase using one of the methods given on the mark scheme. Occasionally, a mark was lost when candidates gave the answer as the percentage multiplier 107.7, instead of the percentage increase 7.7%. A common incorrect answer was when candidates had instead subtracted the two CPI's (111.6 – 103.6) to give an answer of 8%.

There was a general lack of familiarity among candidates with weighted index numbers as indicated by part (c). Candidates were given the weighting for non-alcoholic beverages and the price index for November 2022 and had to calculate the price index for mineral waters, soft drinks and juices. This part of the question was often left blank or a mix of incorrect calculations involving the numbers given in the question. Most incorrect responses were finding 115.6 as a percentage of 124.2, subtracting the two given values or calculating $\frac{115.6 \times 2}{7}$ and not involving the 124.2. Usually, those candidates that could set up the correct equation were then successful in getting the correct answer of 126.7.

Part (d) asked candidates to compare the relative importance of non – alcoholic beverages for 2021 and 2022. The correct answer of coffee, tea and cocoa importance has reduced or more important in 2021 was often seen. Common incorrect answers were candidates also stating that mineral

water, soft drinks and juices had also reduced. Many candidates also did not understand the question and would refer to the quantity of the items sold or had incorrectly read the question and instead compared the weighing of the two different groups instead of comparing the two years.

The final part of this question required the candidates to calculate the geometric mean for the UK and proved to be very challenging. Although many candidates understood the process for calculating a geometric mean, only a small minority knew to convert to the overall percentages (or multipliers) prior to performing the calculation.

Candidates who found the geometric mean of the values, as presented in the table, were able to gain partial credit for knowledge of geometric means. The most common incorrect method was calculation of the arithmetic mean which no credit was given for. Regardless of the method used, candidates were good at interpreting the value of their answer to make a comparison between the average annual inflation for Germany from 2017 to 2021 with the UK.

Question 8

This question presented candidates with a plan for an investigation including a description of how Roxann planned to collect data and how she planned to process and present the data. The candidates were asked to discuss whether these plans were appropriate. It was pleasing to note that the vast majority of candidates made a significant attempt to answer this extended response question, with very few blank responses.

Many candidates were able to gain some marks, but candidates were limited with their marks due to giving statements which were too simple and not justifying their reasoning clearly or didn't realise that their response was supposed to check how well the plan provided would be able to check for a normal distribution. It appeared that some candidates did not realise that 6 marks required 6 statements and, in a small number of cases, did not recognise that they should comment on both the collection of data and the processing & presenting of data.

Common correct observations from candidates for the data collection plan included identifying that it was appropriate to take a census as this would be representative, no control over extraneous variables and not appropriate for runners to record own data as they may lie or record their heart rate incorrectly. They generally did not identify that the data may need to be cleaned or that runners would also need to include their gender on the spreadsheet. Many candidates referred to there being unequal number of males and females or that taking a census would be time consuming and instead a sample would be better which did not receive any marks.

For presenting and processing data, correct observations included that the standard deviation would be helpful for deciding if normal was appropriate, a box plot would allow them to see skew; averages would be equal and separating male and female would be sensible as there may be a difference between the genders. Many candidates commented that box plots and calculating all three averages was appropriate but instead described what

was seen on a boxplot or the advantages of calculating each average instead of why it was appropriate to help deciding if heart rates would be normally distributed. It was rare to see candidates suggesting to draw a histogram to show the shape of the distribution and would instead make reference to drawing curves but not naming them. A common incorrect approach was to describe presenting data in scatter graphs and calculating spearman's rank correlation coefficient and candidates looking for correlation instead of the data fitting a normal distribution.

Question 9

In this question, candidates were presented with Pearson's product moment correlation coefficient and eight regression equations between the length of particular bones and the height of the skeleton for some males and females for ancient Egyptian bones.

Part (a) was generally well answered, candidates were able to recognise that the tibia should be used to estimate the height of an ancient Egyptian because in both male and female bones, it had the highest Pearson's product moment correlation coefficient. A small minority of candidates instead referred to their knowledge of the human body and chose the femur as this is the largest bone in the body and did not use the data given to them in the table. Others incorrectly referred to the bone with the highest gradient in the table or chose the bone that had regression equations with the most similar gradients for males and females.

Part (b) was poorly answered with many candidates unable to interpret the figure 2.594 in the regression equation for the humerus. Many just said that it was the gradient or instead described how they would substitute a value into the humerus equation to estimate the height e.g. 'multiply by 2.594 and then add 83.85'. Others who understood the aim of this question got confused between height and length of the bone and incorrectly stated 'for every 1cm of height, the length of the humerus increases by 2.594cm'.

Part (c) was also not answered well. Many candidates struggled to understand that the gradient given in the equation related to the rate of increase, therefore were unable to suitably explain and compare the relationships between bone length and height for the different bones and between males and females. Where candidates were able to give a partially correct answer, this was often just the first mark for identifying as all bones increased in length the height increases. Many candidates just listed which bones had the highest/lowest gradients and intercepts or would state that males were taller than females.

In part (d), some candidates were able to recognise that using the y-intercept would be unsuitable as this would relate to a bone length of 0cm. Although, for those responses, some candidates mixed up bone length and height and stated that it referred to a height of 0cm. Many candidates did not recognise that the y-intercept was for a bone length of 0cm and would

say that it was appropriate and gave many different reasons as to why, often that it could be used to compare males and female heights.

A significant number of candidates left part (e) of this question blank. Where candidates did attempt this question, the most common correct response was stating that it was appropriate because the regression equations were based on ancient Egyptian skeletons and also the bones were. Others instead incorrectly said it was appropriate because the correlation was high or inappropriate because it was only an estimate or due to extrapolation.

Question 10

Overall, this question on the Petersen capture recapture method was well answered, particularly part (a).

Part (a) was extremely well answered with the majority of students correctly calculating 840. It was rare to see an incorrect answer, or an answer not given as an integer.

In part (b) candidates were asked to discuss the validity and reliability of the estimate found in part (a). The main issue here is that often candidates confused validity and reliability, but credit was still given for this. Many candidates achieved at least two marks with the most popular response being 'unreliable, as the population may have changed due to births/ deaths/ migration.' There were lots of incorrect references to tags coming off or insufficient time for the whales to mix back into the population, showing that candidates had not read the question and were instead just rhyming off conditions that they had been taught about the Petersen capture recapture method.

Question 11

The final question on the paper was a Venn diagram and was a very accessible question with the majority of candidates scoring marks.

In part (a), the vast majority of candidates extracted the correct value from the Venn diagram and gave a correct answer of $\frac{27}{50}$.

In part (b), a large proportion of candidates were able to understand the conditional probability notation and gave the correct answer of $\frac{12}{32}$, although, it was common to see an incorrect answer of $\frac{12}{50}$ where candidates had instead given $P(A \cup B)$. Where marks were not awarded in part (a) and (b), this was due to answers being given as integers instead of probabilities.

In part (c), candidates were asked to conclude if A and B are independent events and this was the least successful part of this question. The majority

of candidates instead gave a description of mutually exclusive events instead of independent stating 'no, because you can have a science fiction book that is also an audio book' or 'no, because the circles overlap'. Those scoring in part (c) were generally divided between using the two methods on the mark scheme.

In the final part of this question, candidates were given some probabilities and had to find $P(C \cup D)$. For those candidates that were successful on this part of the question, they drew a Venn diagram using the probabilities given, thus realising that they had to subtract 0.24 only once. Those who were not successful tended to add together the probability of events C and D and then subtract 0.24 twice.

Summary

Based on the performance of candidates on this paper, candidates should:

- Ensure that correct statistical language is used throughout when making comparisons, just stating figures, or using the word 'whereas' is not a comparison.
- Practice calculating relative risk and interpreting relative risk.
- Practise analysing plans for data collection and the subsequent diagrams and calculations together with giving statistical reasons or against the approaches suggested.
- Practice calculating geometric mean for a range of different contexts.
- Develop their understanding of independent events.
- Develop their understanding of regression equations and the interpretation of the gradient of these.

