Pearson Edexcel

Examiners' Report
Principal Examiner Feedback

November 2020

Pearson Edexcel GCSE (9-1)

In Statistics (1ST0)

Higher Paper 1H

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

Grade Boundaries

Grade boundaries for all papers can be found on the website at:
https://qualifications.pearson.com/en/support/support-topics/results-certification/grade-boundaries.html

November 2020
Publications Code 1ST0_1H_2011_ER

GCSE (9-1) Statistics – 1ST0

Principal Examiner Feedback – Higher Paper 1

Introduction

General comments

Most students responded to the challenges within this paper well. They were generally confident at completing calculations and diagrams and demonstrated good statistical understanding when asked to interpret these. Students found questions requiring evaluation of approaches or techniques more slightly more challenging.

Students should be reminded that when drawing any form of graph or diagram they should use a ruler and a sharp pencil to ensure accuracy. They should also be encouraged 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

In part (a) the vast majority of students were able to correctly identify the type of data collected by the question.

Most students were able to give a partial explanation of how to use a list of random numbers to select a simple random sample of 50 students as required in part (b). A small number of students referred to drawing names from a hat or generating random numbers in some way, this was not required as the question specifically asked how a list of random numbers could be used to select a simple random sample of 50 students. Students were generally able to describe numbering the list of students and including those that match to the numbers selected in the sample. Many students did not refer to excluding numbers outside the required range or picking again if a repeated number was seen.

Question 2

Almost all students were able to correctly interpret the choropleth map to identify the percentage change in the number of police officers in region 17 (part a). The majority of students were also able to correctly identify the number of regions in which the number of police officers increased (part b).

In part (c) of the question students were asked to explain why the interpretation given might not be correct. This caused more difficulty than the previous parts of the question, with many students making reference to where the percentage might lie within the interval given or adding up the percentages rather than identifying that the conclusion related to the number of police officers rather than the percentages and that the numbers of officers for each region was unknown.

The majority of candidates were able to correctly identify that the diagram was a choropleth map.

Question 3

This question was well answered by the majority of students. Most were able to read from the composite bar chart to find an estimate for how many more new car registrations there were in the UK than in Italy (part a). In part (b) most students correctly identified that part (a) was an estimate because exact values could not be read from the graph or that the values from the graph were given in thousands or tens of thousands. The small proportion of incorrect responses seen in part (b) referred to the possibility of a different type of fuel or stated that it was an estimate because it wasn't precise without giving any further explanation.

Part (c) was also well answered with most students able to give three correct comparisons between the number of new car registrations in the UK with the number of new car registrations in Italy for the different fuel types. A minority of students only gave comparisons for one or two of the fuel types. In a small number of cases students merely stated values without comparison or made comparisons between different fuel type frequency in the UK.

The majority of students were able to correctly answer part (d) of this question identifying that only quarter 1 was given and not data for the rest of the quarters.

Question 4

In this question most students demonstrated understanding of tree diagrams and their use to calculate probabilities. In part (a) the tree diagram was usually fully correctly completed, similarly most students were able to show the required probability in part (b).

It was pleasing to see that most students were able to demonstrate that the claim in part (c) was correct by finding the required probability using the tree diagram and comparing to 6%. A minority of candidates found the probability of fake followed by genuine but did not double to allow for the two possible orderings. A small number of responses calculated a correct probability but did not give a conclusion.

Question 5

Students were generally able to interpret the table correctly in part (a) and as part of calculations in part (b). The majority of students were also able to explain why 'Cleanliness of the inside' had the most reliable percentages, although a minority incorrectly said that this was because this category had the highest percentage for 'satisfied or good'.

Question 6

It was pleasing to see that most students were able to find the 90^{th} percentile from the cumulative frequency graph (part (a)(i)). A minority of students found 27, but thought that this was the 90^{th} percentile rather than identifying this as indicating that they needed the 27^{th} value from the graph. Students found interpreting the 90^{th} percentile in context (part (a)(ii)) more challenging, with a common error being referring to the temperature on the 27^{th} of June being equal to the value obtained in (i).

In part (b), a majority of students were able to use information from the cumulative frequency graph in order to demonstrate that the greatest maximum daily temperature in June is not an outlier. Some students were aware of the calculation required, but did not find the lower quartile and upper quartile correctly. Others found the quartiles, but did not use these correctly to demonstrate that June was not an outlier.

Part (c) of the question required students to draw a box plot. There were some fully correct responses, but it was not uncommon to see errors in plotting of some of the values and in some cases students appeared not to know how to find the values required from the cumulative frequency graph.

For part (d) of the question students were required to compare the distributions of maximum daily temperatures in June for the two locations. Most students were able to make two correct comparisons, often of median and interquartile range. A correct comparison of skew was less common. A minority of candidates compared lowest values or highest values etc which was not sufficient.

Question 7

Part (a) of this question was generally answered well, with the majority of students able to identify that a hypothesis should be a statement rather than a question.

In part (b) students were asked to explain why stratifying would be an appropriate method of sampling. This was less well answered with some incorrectly referring

to this allowing comparison between genders and between ages which were the strata being used.

Students found part (c) of the question challenging. In part (c) they were asked whether the data collection sheet given was appropriate, however it was common to see answers that referred to the sampling approach rather than to the data collection sheet. Where students did comment on the data collection sheet, they often referred to the problems such as too few options, students may use more than one option or that students' ages were not recorded. It was uncommon to see comments identifying that the data collection sheet would make it easy to analyse responses or put the results into graphs.

In part (d) students were asked to explain why comparative pie charts would be appropriate to present the results of the two investigations. Students were expected to indicate that the number of students at the university would be represented by the size of the pie charts. Some students were able to explain this, but others made reference to the pie charts showing proportion without acknowledging the different sizes of the samples which was not sufficient to justify the use of comparative pie charts.

Part (e) of the question required students to calculate the diameter to use for the second of the two comparative pie charts. Whilst some students were clearly familiar with the appropriate calculation and correctly found the required diameter, others appeared not to know how to start and 15cm was a common incorrect answer which came from using the ratio 12000 : 18000 as the ratio for the diameters.

In part (f) students were asked to discuss things that should have been considered in planning the investigations which would help improve the reliability and validity of their comparisons. This proved challenging for students. Students were expected to identify factors such as ensuring the sample size was not too small, using similar criteria in selecting students, asking questions in the same way or asking the same question, doing the investigation at similar times or in similar venues. Often students gave responses that did not focus on the comparison aspect which was required. Some students gave very vague responses which did not have sufficient detail. A common incorrect answer was stating that they should ask the same number of people. Another common answer which did not gain credit was suggesting use of random response which was not relevant in this case as there was not a personal or sensitive question being asked.

Question 8

In part (a) of this question required students to show the calculation for a given standardised score. This was answered well by most candidates.

Parts (b) and (c) required students to compare the results that a student had in tests in different subjects. This is a common application of standardised scores and most students were able to answer the question correctly. In part (b) a small number of students did not recognise that the standardised scores were relevant and instead compared the mean marks or the standard deviations. In a couple of instances students appeared to have confused standardised scores with correlation coefficients as references to closer to 1 or closer to -1 were seen.

Question 9

Students were generally able to calculate the number of pigeons in the sample that were tagged by use of the Petersen capture recapture formula (part a). The majority of students were also able to identify some of the assumptions used for this method in part (b), although some also included reference to sample size. There were a significant number of students who omitted the conclusion or who indicated both reliable and not reliable.

Question 10

Calculation of a chain base index number, in part (a) of this question, was generally done correctly by students. Where incorrect answers were seen these were based upon attempting to find patterns within the numbers in the table or subtracting.

Most students were able to identify that the working and conclusion given in part (b) of the question was not correct. A majority were also able to go on to give a clear and correct reason why this was. Where incorrect answers were seen in (b) these often followed incorrect answers in (a).

The majority of students were able to identify that the geometric mean of 102.18 related to a 2.18% increase, although some did not go on to give a full correct answer as they omitted the 'per year' from their interpretation.

Question 11

Part (a) was generally correctly answered by students who identified that the trend of visitors was upwards or increasing. Where incorrect answers were seen these often described the variations step by step – indicating increasing then decreasing for example – or referred to positive correlation which was not an acceptable answer.

Students were also able to identify the quarter each year which the museum had the fewest visitors (part b). Where incorrect responses were seen these generally referred to a single quarter of a single year.

In part (c) of the question students were asked to interpret the gradient of a trend line in context. Many students found this challenging and only a minority could answer this correctly. Common errors included omitting the contextual element of the interpretation, omitting reference to per quarter or including a time frame other than one quarter, or making reference to multiplication by 1.4 rather than increase by 1400 visitors.

Part (d) of the question asked about the validity of a prediction using the trend line and mean seasonal variation to predict outside the range of data. Whilst some students were able to identify that this was extrapolation and so the prediction might not be correct (or similar), it was also common to see incorrect answers.

In part (e) of the question students needed to use the trend line and mean seasonal variation to predict the number of visitors before comparing to the actual data. Students found this challenging and often gave just the trendline value or averaged the values for the quarter 2's for previous years. Where students did make an attempt at the calculation required it was not uncommon to see errors in reading the scale or omission of the numerical comparison to the given value.

Question 12

It was pleasing to see that the majority of students were able to correctly identify Spearman's rank correlation coefficient as the required calculation, could perform this actually and interpreted this to determine the extent of agreement between the judge and the Mayor. There were, however, a significant proportion of students who could identify that Spearman's rank correlation coefficient was required, but made errors in applying the formula. Only a minority of students could not identify the required technique.

Question 13

Students found this question relatively challenging. In part (a) students were asked to consider the conditions required for a binomial distribution to be a suitable model. Many students struggled to give any of the required conditions. Where conditions were identified these were often not given in context as was required for full marks to be awarded.

Part (b) of the question required identifying a simple probability and this was generally correctly answered.

In part (c) of the question students were expected to use the binomial distribution to find a probability. This was not well answered. Common incorrect answers did not use the binomial distribution. Some students attempted to use the binomial distribution but made errors with the powers of 0.25 and 0.75 or in finding the binomial coefficient.

Summary

Based on their performance on this paper, students should:

- Practise writing clear explanations, bearing in mind exactly what is asked in the question and what evidence you should give to support your answer.
- Practise interpreting statistical calculations in the context of the question.
- Develop skills in evaluating approaches to statistical methods.
- Develop skills in planning for data collection and evaluation of proposed data collection approaches.
- Practise use of trend lines and mean seasonal variation to make predictions for time series graphs.
- Practise calculation of Spearman's rank correlation coefficient.
- Practise calculating probabilities using the binomial distribution.