edexcel

Principal Examiner Feedback

June 2011

GCSE Statistics (2ST01)

Unit 1: 5ST1H_01 Higher

June 2011
Publications Code UG028833

1. PRINCIPAL EXAMINER'S REPORT – HIGHER PAPER 1

1.1. GENERAL COMMENTS

1.1.1. This was the first examination for the new specification and the style of the question paper was slightly different to that of specification 1389

1.1.2. There was a greater requirement for interpretation of various statistics and diagrams.

1.1.3. Candidates should be encouraged to show their working as some may have picked up more credit when their answers were incorrect. In some cases it was evident that correct values were extracted from the question but then incorrectly added etc, suggesting lack of a calculator. There was a general improvement from the previous specification in the standard of diagrams; candidates should take care however with reading correctly the scale on axes. They should also be encouraged to use a ruler when drawing box plots, histograms, lines of best fit, etc.

1.1.4. Candidates tackled the interpretation quite well. On occasion they forgot that this was a statistics paper and used none statistical terms.

1.1.5. Candidates seemed to have no difficulty in finishing the paper in the time allotted.

1.1.6. Poor handwriting continues to be a problem, as is the habit of not showing working. The latter meant a candidate lost marks e.g. writing down an answer of 3.94 when one decimal place (3.9) is required. If no working was shown all marks (4) were usually lost.

1.1.7. In questions that are marked *, which denotes a QWC question, candidates are expected to use correct statistical terms.

1.2. REPORT ON INDIVIDUAL QUESTIONS

1.2.1. Question 1

This question proved to be very accessible to most candidates. It was done very well. Most candidates were able to write down at least 2 of the things wrong with the graph. Some had difficulty explaining that it was the lack of scale that was wrong with the vertical axis. A significant number of candidates did not refer directly to 3D but explained that the graph was at an angle. Another fairly common error was for candidates to criticise labels on the horizontal axis.

1.2.2. Question 2

Generally this question was done well. A few candidates left out part (a), those who did do it invariably got it right.

In part (b) most candidates were able to write down both negative correlation and a practical interpretation of the problem. Candidates should be advised to describe the correlation fully, i.e. positive correlation and not just positive (on its own).

In part (c) it was pleasing to see so many candidates draw their line of best fit through the mean point.

In part (d) the majority of candidates attempted to give a practical reason that may have been correct in intention but was ambiguous and did not get a mark. The general reason (extrapolation) was relatively rare.

1.2.3. Question 3

This question was generally done well, with most candidates showing their working in parts (c) and (d).

In part (c) a few candidates continued working from their correct answer of 60.9 to get an incorrect answer or used the 100 – (19.6 + 9.1) approach.

Many candidates were able to give a correct reason in part (e), but a significant number of candidates thought that the sample was incomplete, rather than an issue of non-response or thought that the reason was a rounding error (the difference was too great for this to be correct).

1.2.4. Question 4

Part (a) was generally done well. The most common incorrect answer here by far was to state their hypothesis as a question.

Part (b) was done well with most candidates stating their reasons separately in each space provided, but some put both of their reasons in the same space- typically 'cheaper and quicker'. Many did not realise that the subject of their comment was the sample. Candidates should be advised to write their answers in complete sentences at all times, e.g. 'the sample is cheaper', 'the census takes more time', etc.

Part (c) was generally done poorly; most candidates either gave a sampling method as their answer or suggested a small number of the students at the university.

Few candidates were able to give an 'in context' response to part (e). A common error was to confuse control group with pilot study. Some candidates recognised that a control group is used to compare but thought that gender would be compared.

1.2.5. Question 5

A common error in part (a) was a grand total of 188.

In part (b) most candidates were able to identify Paris/Venice as the required holiday, but many were unable to give a correct (unambiguous) reason for this. A common incorrect answer here was 'largest in the table'. There were also a number of candidates who discussed the relative attractions of the cities.

Part (d) was the least well done part of this question. A very common incorrect answer here was 32/94.

1.2.6. Question 6

Part (a) was quite well answered.

Part (b) was also well answered with many candidates referring to it being an open question or that it assumes use of the facilities. Some simply said it was a leading question which wasn't enough of an explanation, however most who said it was a leading question went on to elaborate why and therefore obtained the mark. A small number of students were confused by the wording of the question and comments were seen such as "there is only one way to use the facilities so it is badly worded".

Many candidates were able to do Part (c) but some gave no option boxes or gave option boxes that were overlapping or non-inclusive.

1.2.7. Question 7

Many candidates got correct answers to parts (a) and (b). The minority of candidates who did not start by writing the **15** values out in order often lost marks in these two parts.

In part (c) quite a few candidates were able to get 2 – 3 marks for comparing the distributions, but only the best were able to score full marks. When comparing distributions candidates should compare a measure of centrality and a measure of spread, in this case median and either the range or IQR; only after that should they look at other comparisons such as skew.

A significant number of candidates knew that they were expected to compare the skews of the distributions but were unable to do this using the appropriate language, e.g. they wrote down such expressions as 'the median is closer to the right…'. Unfortunately many candidates used the word spread rather than IQR or range and mean or medium rather than median. This question did require candidates to use correct statistical language. A number of candidates did not at any point say 'there were more shoes sold during the sale'.

1.2.8. Question 8

In part (a) many candidates did not understand what they were being asked. Many of these concentrated on the practical difficulties of collecting the information, e.g. people not answering the phone, rather than the theoretical considerations of representative-ness. A significant number of candidates thought that the sample was representative because ‘a lot of people are being asked’.

In part (b) candidates demonstrated their usual difficulties in defining how to take a random sample with the added complication that they were also expected to explain the correspondence between the chosen numbers and the people used in the sample. Candidates should be advised to look at the number of marks available for a question. The number of marks determines the number of separate things required in the answer.

Part (c) was generally done well. Many candidates were able to show their understanding of this topic. A minority of candidates decided that it was necessary to write a short essay to attain the 3 marks available, often failing to answer the question and commenting on how to construct a questionnaire; a sizeable percentage of these achieved only 1 or 0 marks as a consequence. Many of the successful answers consisted of 2 or 3 sentences identifying the considerations of cost, time, response rate and veracity.

1.2.9. Question 9

Candidates were generally successful with part (a) of the question. Candidates should be advised to show some working when the question is worth more than one mark. Incorrect answers here generally lost both marks in (a).

In part (b) candidates filled the empty columns with a variety of things (including class width, frequency density, cumulative frequency, etc) though a good number were able to identify the class mid-point. A common mistake was to see the mid-point for $0.2 < h \leq 0.5$ given as 0.3. Division by 8 was another common mistake which lost candidates the final two marks of the question. It was not uncommon for the correct *fx* values to be totally ignored and a different approach subsequently adopted.

It was pleasing to see that many candidates were able to give their answers to the appropriate degree of accuracy as stated in the question.

1.2.10. Question 10

Parts (a) and (b) were generally done well.

Part (c) was not done well. Few candidates could explain that the increase in numbers was shown by the larger area. The vast majority tried to explain the difference in sociological terms.

1.2.11. Question 11

Parts of this question were done well. In part (a) it was pleasing to see so many candidates able to calculate the 4-point moving averages correctly - candidates should be advised to show their calculations clearly here. Plotting the moving averages and drawing in the trend line was also generally well done, although a minority of candidates joined up the points ‘dot to dot’ style. The majority of candidates were able to describe the trend and obtained the first mark usually saying the trend was positive or increasing. Fewer candidates answered the question on interpreting the trend. The most common errors were to not include an interpretive answer in context, or to not include a time frame for example “more motorcycles are registered” was often seen.

In part (d) a minority of candidates incorrectly wrote 2007, Q2.

Parts (e) and (f) were only done well by the most able candidates. The most common wrong answer to part (e) was 43 arising from students averaging the three Q2 values (43.5, 41.4 and 44.2). The students with the most success used the three values (43.5, 41.4 and 44.2) from the table, subtracting their trend line values from these before averaging. Some students were reading values from the graph inaccurately, however these students were often still within tolerance and obtained the marks. Another incorrect method seen was to work out the average difference between quarter 2 and quarter 1. Those candidates who got part (e) correct usually got part (f) correct as well.

1.2.12. Question 12

In part (a) most candidates were able to correctly rank the numbers and start to use the formula. It is perhaps surprising that quite a number of candidates were unable to copy the formula correctly from the formula page. A common error in calculating the d^2 values was usually due to dealing incorrectly with the squaring of negative integers.

In part (b), of those candidates with a correlation coefficient to comment upon, the majority were able to identify their answer as positive, negative or no correlation. A number of candidates lost a mark for only describing the strength of the correlation, e.g. weak correlation. Only a minority were able to place their Spearman value in the context of the question. A surprising number of candidates described the relationship between the variables as a positive correlation (which we accepted) despite the low value of the coefficient.

1.2.13. Question 13

Generally this question was done quite well. In part (a) many candidates were able to both work out the correct heights of the bars and draw them in the diagram with the correct widths. Only the weaker candidates were tempted to complete the diagram as a bar chart.

In part (b) a significant number of candidates simply gave the answer (15) without showing any of their working- this is an all or nothing approach which should be discouraged.

1.2.14. Question 14

Only the best candidates were able to score full marks in this question. In part (a) most candidates managed to put in 0.9 and 0.2 correctly but few put in 0.05 and 0.95 correctly – values of 0.8 and 0.2 were often seen repeated. A number of candidates showed their lack of understanding on how to complete a tree diagram by putting values in each 'pair' that added to 1.

Part (b), a question on conditional probability was badly done and was clearly beyond the capability of most candidates. An incorrect answer of 0.125 was common. In part (c)(i) few candidates recognised the correct distribution – normal, bimodal or a sampling method were common incorrect answers.

In part (c)(ii) many candidates were able to give only one of the required reasons, usually the 'probability of getting the allergy and not getting the allergy'. In part (c)(iii) many candidates identified correctly the need to calculate $10p^3q^2$, but most were unable to calculate this correctly. A surprising number of candidates used values of p and q which did not total 1, often with values of p and/or q each in excess of 1, e.g. p=2 and q=3.

1.2.15. Question 15

Part (a) was done well by many candidates. Many stated the leaflet had been a success and gave an acceptable reason interpreted from the graph. Some candidates were confused about the 24%, thinking that this was referring to the total reduction required for the success.

Part (b) was done less well with often only one of the required reasons given. 'The figures shown for women are less than the figures shown for men' was commonly omitted. Many candidates used figures from the chart to back up their reasons. It was fortunate for many that generally the accuracy of doing this was not an issue here.

1.3 GRADE BOUNDARIES

Grade boundaries for this, and all other papers, can be found on the website on this link:
http://www.edexcel.com/iwantto/Pages/grade-boundaries.aspx

Further copies of this publication are available from
Edexcel Publications, Adamsway, Mansfield, Notts, NG18 4FN

Telephone 01623 467467
Fax 01623 450481
Email publication.orders@edexcel.com
Order Code UG028833 June 2011

For more information on Edexcel qualifications, please visit
www.edexcel.com/quals

Pearson Education Limited. Registered company number 872828
with its registered office at Edinburgh Gate, Harlow, Essex CM20 2JE

Llywodraeth Cynulliad Cymru
Welsh Assembly Government

Rewarding Learning