Version: 1.0

General Certificate Secondary of Education
June 2012

Statistics 43101H

(Specification 4310)

Unit 1: Higher

Report on the Examination

Further copies of this Report are available from: **aqa.org.uk**

Set and published by the Assessment and Qualifications Alliance.

The Assessment and Qualifications Alliance (AQA) is a company limited by guarantee registered in England and Wales (company number 3644723) and a registered charity (registered charity number 1073334).
Registered address: AQA, Devas Street, Manchester M15 6EX

.

Unit 1: Higher Tier

General

The examination proved to be accessible for students of all abilities.

There was a deliberate decision taken to have more questions on the paper that could be accessed by all Tier H students but at the same time continue to maintain the subject rigour and integrity expected at this Level. This change in emphasis was reflected in the median and quartiles showing an increase of around 12 marks compared to the 2011 distribution.

There were few questions not attempted on the paper and there was a significant improvement in both the presentation of answers and basic understanding of statistical concepts and methods. Centres appeared to be reacting well to the requirements of the new specification with students better able to respond to the emphasis now placed on questions relating to data interpretation and inferential Statistics.

Most students achieved high marks on the common questions between tiers and the standard of answers to questions on probability and index numbers showed significant improvements. However, certain parts of the specification remain problematic. These include, the Normal distribution, control charts and regression analysis. Additionally students had difficulty with questions on measures of skewness, correlation analysis and sampling principles.

Some students need to take more care with the graphing of data and levels of accuracy in their answers. Calculators should be used, where appropriate, throughout the paper, but particularly with probability, where incorrect answers resulted from basic errors of addition and / or multiplication of decimals.

Topics that were well done included:

- histograms including use of population pyramids
- probability based on a completed tree diagram
- index numbers
- interpretation of cumulative frequency graphs.

Topics which students found difficult included:

- normal distribution, with specific reference to standard limits
- regression analysis
- calculation of the standard deviation using given summary measures
- interpretation of measures of skewness.

Question 1

Part (a) was generally poorly done. Quite a few students clearly thought that the low numbers were given to Critical because they were the most important or life threatening type of illness. Of those who seemed to grasp the idea only a small number were able to explain clearly that 00 - 49 was 50 numbers and one half of all the 2-digit numbers; and 40 was half of the 80 patients. Many said it was “even” or about $\frac{1}{2}$.

In (b) parts (i) and (ii) were very well answered with (iii) producing a range of answers many of which were correct.

Question 2

All parts were well attempted. A few failed to use both of the over 70 categories in part (b) but the interpretation was generally correct in part (d).

Question 3

A considerable number of students did not recognise this terminology and many students simply ‘guessed’ the answers.

Question 4

There were many theories about the reason in part (d) - often involving crime rates or population differences. Those who scored tended to focus on areas such as size and / or quality of the police force in Scotland. Part (e) proved to be a difficult question for many. Very few were able to answer successfully with the vast majority suggesting "rounding errors", "crimes that weren't reported" or that "some crimes were not detected."

Question 5

Most students gave a correct response in part(a). However, two common incorrect answers were "it is the first column" or "they are all the same". In part (b) most understood what the figures were showing and simply stated "it stayed the same". In part (d) a significant number seemed to think there was a physical interpretation to the weighting. The calculations were generally well done in part (e), although a sizeable group failed to understand the idea of a base number and used 3% of 630 in part (i) which leads to the incorrect answer of 648·90 or added 8% onto 360 to give 680·40. In part (ii) similar confusion often led to a variety of incorrect answers although a small number did manage to recover to obtain marks on follow through.

Question 6

Part(a) (i) was usually correct but in part (ii) very few seemed to realise that a correlation coefficient of such a small magnitude meant no correlation between the variables. Most incorrectly tried to describe it as a very small negative correlation, or similar.

Question 7

Almost all students named the sampling method correctly and managed to show how the numbers had been obtained. Part (b) was quite well answered but some students attempted to use a selection based on names / numbers from 'a hat' and there were frequent vague references to random number generators without any clarity about what these are. Quite a few seemed to think they put numbers into a random number generator. In part (c) most realised that first class passengers would be missed out but few spotted the shortfall in numbers using this method.

Question 8

Parts (a) and (b) were poorly done. Some students plotted the point (9, 9·5) incorrectly and others were unable to draw a line with a gradient of $-\frac{1}{2}$. The interpretation of the graph in (b) (ii) was poorly attempted with the vast majority trying to describe the problem in terms of extrapolation, variability of lines or some other issue relating to the drawing of the graph. Very few realised the meaning of the point in terms of days lost at work. In part (c) whilst many used their graph to read off the values there were quite a few, possibly the more able mathematically, who used their equation, even though some of the answers thus obtained were quite unrealistic. This was one question where a seemingly correct answer could be found from a wrong method as there were several who used just the two given points at (17, 2) and (22, 6) and showed that the midpoint was (19.5, 4) so the value at 20 must be about 4. There was a lot of confusion in part (d). More spotted the negative value than the extrapolation problem, although quite a few stated that the graph did not go that far, rather than the values not going that far.

Question 9

Most parts were well done. There was however, some confusion in part (e) where many gave the wrong port as the answer. In part (b) there was a fairly common incorrect method leading to a right answer. Using 60 minutes in an hour quite a few gave the 1st decile as $60 \div 10 = 6$, the 9th decile as $9 \times 6 = 54$ then found the difference, $54 - 6 = 48$.

Question 10

This aspect of data presentation seemed well understood and a large majority got the correct answer. The few who did not score usually just used the heights of the bars instead of the areas of the rectangles. Using the given formula for standard deviation caused problems for many students. Some substituted $\sum fx$ into the formula instead of $\sum f$. Another sizeable minority used 2 6 instead of 2 6^2 thus obtaining a fairly common incorrect answer of 5·1.

In part (c) most were able to interpret the averages and realised that men had lost more weight but few were confident in their interpretation of standard deviation. There were however, a few who failed to interpret the figures at all and simply quoted numbers. When calculating skewness a significant number were unable to cope with the order of operations but generally most were able to use their previous answers to calculate suitable measures.

Question 11

This question was not well done. In part (a) the incorrect responses of $157 \pm \sigma$ and $157 \pm 2\sigma$ (rather than $157 \pm 3\sigma$) were common, leading to the incorrect answers of 153·5 – 160·5 or 150 – 164. A few students did use the more accurate version of $157 \pm 3{\cdot}09\sigma$ that is quoted in the specification.

Some students gave answers which were approximately correct but with no working shown indicating how these values had been achieved, it was not possible to give any credit. In part (b), of those students who did realise that the stated weight was 2 standard deviations away and knew that approximately 95% of the data lay symmetrically between these values, unfortunately many did not divide by 2. A common error was to use 150 / 157 as a percentage giving 95·5% which was seen several times and was close to the correct value for the 2σ difference.

In part (c)(i) there was quite a lot of misunderstanding of the meaning of the control charts. Many students picked up the problem with machine A exceeding the action limits and many also spotted that B was satisfactory. However, a number thought that C was also satisfactory as it did not exceed the limits. Many students expected all the values to be exactly on the target weight or thought that any value below the target was unsatisfactory and any value on or above it was satisfactory. In part (c)(ii) few were able to give a clear indication of the use of a sample range. Many suggested outliers, anomalies or accuracy as the reason for its use. The majority simply gave vague answers such as “to check effectiveness”.

Question 12

A large majority of students completed the tree diagram correctly.

In the calculation parts students generally used decimals but a few converted to fractions for the work. Some examples of premature approximation were seen and occasionally 0·0875 was rounded to 0·08. Part (i) was very well done with the majority getting 2 marks. However, a sizeable number added the probabilities instead of multiplying.

Whilst most students were aware that multiplication was required in part (ii) many failed to double their answer and gave an incorrect answer of 0·0875. A similar problem occurred in part (iii) where many managed to get part marks but failed to take into account all possibilities. In part (c) most realised the need to multiply but a few picked the wrong fraction from the question thus getting $2800 \times 0{\cdot}35$ or $2800 \times 0{\cdot}4$. There was a small number who divided by the probability and gave an answer of 11200.

Mark Range and Award of Grades

Grade boundaries are available on the Results statistics page of the AQA Website.

UMS conversion calculator www.aqa.org.uk/umsconversion