**General Certificate of Education (A-level)
June 2012**

Statistics **SS02**

(Specification 6380)

Statistics 2

Report on the Examination

General

While the level of computational skills remains good, the quality of expression where explanations are required leaves much to be desired. Often answers were not so much incorrect as incomprehensible. Candidates should make sure they frame their answers in terms of the context of the question, rather than simply quoting learnt generalisations.

Question 1

This was a good opening question for the majority of candidates, with many scoring 7 or 8 marks, but the number who incorrectly completed the table was surprising. Part (c) proved difficult for a sizeable fraction of candidates, with 0.5 being a common answer.

Question 2

Most candidates correctly identified *n* and calculated *y*. For many the calculation of x was a simple manipulation of the formula used for *y*, but others treated this as a forecast request. Part (b)(ii) was well answered, although there was some confusion of 'short term' and 'seasonal'. Many seemed to have no knowledge of the term 'seasonal effect' even if they could use it in the forecast required in part (b)(iv). Others ignored the instruction 'using the given regression line' and did calculations from the table or based on a calculator regression line. In giving the answer to part (b)(iv) many omitted the 'millions'. Most candidates correctly recognised that 2012 was too far away from the data for an accurate forecast.

Question 3

The majority of candidates correctly found the missing value in part (a) although a few seemed to treat this as an interpolation exercise. Some were uncertain what was required in part (b) while others erroneously divided their total by 15 or 30. Part (c)(i) was frequently answered as if the question had asked for 'no more than 3', especially by those who seemed to be using a calculator function, but parts (c)(ii) and (c)(iii) were well done. A significant fraction of scripts showed no attempt at parts (d) and (e). Candidates are to be encouraged to check that they have attempted all parts of each question. Part (d)(i) was quite well answered, although some candidates felt that to obtain the mean they must divide 9.6 by 2. Many seemed unaware that 3 months constitute one quarter of a year, but on the whole part (d)(ii) was well answered. Most candidates appreciated that the Poisson distribution would not be appropriate given the additional information and many could explain why not. Others would have scored better if their explanations had been in context rather than generalised statements about the Poisson distribution.

Question 4

Only a small minority could correctly identify cluster sampling, or state clearly that a hypothesis test required a random sample, which this was not. Since both the sample mean and the population mean are talked about, it is important that candidates identify which they are referring to in the hypotheses. Use of μ is the simplest way to do so. The calculations were generally well done, but there was the usual confusion between two- and one-tailed tests. Many found the context statement difficult, often saying "Accept H_0" but following up with "there is not significant evidence that the mean mark **is** 50". Only a small fraction realised that the large sample size allowed the CLT to be used to assume the sample mean would be normally distributed, while for the final part many simply wanted to use 100 instead of 81 without any consideration of what the other 19 might have scored. Others did realise that this knowledge was essential for a valid result.

Question 5

Although most candidates got part (a)(i) correct, inevitably many also treated part (a)(ii) as if it had asked for the percentage **less** than 200 pounds, while others struggled to express 26 as a percentage of 180. The box and whisker plot was often poorly done, with only the end points correctly located. It is disappointing that many candidates at this level do not take the trouble to use a ruler. In part (b)(ii) many recognised that the question asked about the distribution and focused on the range, IQR and degree of skew, while others merely compared point values. A significant proportion compared the means, despite there being no information on these. References to "more of the women are above the median" revealed a lack of understanding of the concept of a median. Part (c) was well answered, most recognising the unrepresentative nature of data from just one doctor's surgery, where health issues might also affect results. Candidates should not assume that "too small a sample" will always be the required answer.

Question 6

Although most candidates could correctly quote, in context, the rules for selecting the sample by using random number tables, a much smaller proportion could actually carry out the task, many including 82 and/or the second 71. Others only moved along by one number, generating 09, 96, 63 etc., failing to recognise that their 'random' numbers were no longer independent. In part (c)(i) 'explain' required a brief description of what a systematic sample was and then **why** this would not necessarily be more representative. Many answers simply reiterated that it would not, and few seemed to appreciate that if the list had been by party this would have generated a stratified sample. Others answered the question they expected to be asked – "Why is a systematic sample not truly random?". Part (c)(ii) was very well answered.

Mark Ranges and Award of Grades

Grade boundaries and cumulative percentage grades are available on the Results statistics page of the AQA Website. UMS conversion calculator www.aqa.org.uk/umsconversion