Examiners' Report/
Principal Examiner Feedback

January 2011

GCE

GCE Statistics S1 (6683) Paper 1

Edexcel Limited. Registered in England and Wales No. 4496750
Registered Office: One90 High Holborn, London WC1V 7BH

A PEARSON COMPANY

Edexcel is one of the leading examining and awarding bodies in the UK and throughout the world. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers.

Through a network of UK and overseas offices, Edexcel's centres receive the support they need to help them deliver their education and training programmes to learners.

For further information, please call our GCE line on 0844 576 0025, our GCSE team on 0844 576 0027, or visit our website at www.edexcel.com.

If you have any subject specific questions about the content of this Examiners' Report that require the help of a subject specialist, you may find our Ask The Expert email service helpful.

Ask The Expert can be accessed online at the following link:

http://www.edexcel.com/Aboutus/contact-us/

January 2011

Publications Code UA026663

Statistics Unit S1
Specification 6683

Introduction

The paper appeared to be of about the right standard and length. Each of the questions were accessible to all candidates but the longer questions 6, 7 and 8 proved more challenging towards the end. Question 3(c) also proved quite discriminating as many students failed to realise that the lower quartile was the point of focus rather than the upper quartile.

There was some good and accurate use of the formulae in the formula booklet and simple applications of the normal distribution are now handled confidently by most candidates.

Report on individual questions

Question 1

This proved to be a friendly starter for most candidates with many scoring all 5 marks in part (a) and (b). Most errors here were arithmetic such as writing S_{ll} = 596.666… rather than 569.666… or accuracy problems in part (b) where an answer of 0.57 was often seen, rather than the 3sf accuracy that we look for. A minority still have difficulty in using the printed formulae and $S_{ll} = 327754.5 - \left(\frac{4027}{50}\right)^2$ or $\sum l \times \sum w$ instead of $\sum lw$ in S_{lw} were sometimes seen.

Part (c) caused problems for many candidates who simply wrote "positive correlation" but did not interpret this statement in the context by mentioning that longer salmon usually weigh more. Some candidates tended to "overstate" their conclusion by implying that as a salmon grows it gets longer (not strictly true in this instance as the study was of 50 different salmon not one salmon at 50 different time intervals) and others referred to a proportionate relationship such as "for every cm increase in length the salmon weighs 0.572 kg more". Whilst such indiscretions were overlooked for the single mark on this occasion, these examples should provide useful points of discussion for teachers with future cohorts of students.

Question 2

Again the majority of the candidates encountered few problems here and many scored 3 or 4 marks, although a number in part (a) found both means and then simply added them and divided by 2 rather than taking the weighted average. Most chose to find the sum of all 28 records and then simply divided this by 28. A few candidates misinterpreted the 84.6 as the mean for the 21 days despite what should have been an obvious discrepancy with Keith's data.

In part (b) many realised that the changes would have no effect on the mean but sometimes they failed to give any numerical values to support this claim. There were many excellent answers though showing clearly that 9.4 + 0.5 = 4.9 + 5.0 or some other suitable calculation.

Question 3

There were many fully correct box plots in part (a) that gained full marks but also a number of almost correct plots with no supporting working that scored very few marks. Those who did show their calculations for determining the outliers usually got the 24.5 value correct but a large number had 3.5 rather than – 3.5 as their lower limit. Some candidates drew two upper whiskers one ending at 20 (the next highest non-outlier data point) and one ending at 24.5 (the outlier limit). A correct answer should only have one whisker and it appears that some candidates had been copying the practice on the mark scheme of illustrating both alternatives for the benefit of the examiners. It should be remembered that the published mark schemes are <u>not</u> model solutions and they should not be offered to candidates as such.

In part (b) most candidates stated that $Q_2 - Q_1 > Q_3 - Q_2$ or something equivalent though some then claimed that the skewness was positive but this was usually answered well.

Part (c) caught many candidates out with a large number agreeing with the company's claim because Q_3 = £14 000 which is greater than £10 000. Those who did appreciate that the lower quartile was the significant figure to be considering often gave excellent answers but the success rate here was lower than expected.

Question 4

Part (a) was answered well by many but a number still have difficulties in determining which variable corresponds to *y* and which to *x* and then using the formulae given in the booklet. The question clearly stated that the regression line of *p* on *v* was required and this was further emphasised by giving the form as $p = a + bv$ but despite this a number calculated 1.688/1.168 for *b* and some used $a = 4.42 - 3.32b$. There was the usual crop of accuracy errors with candidates failing to work accurately enough to give their final coefficients to the usual 3sf (or better) accuracy. A significant number of candidates simply substituted 85 into their equation and received no marks for part (b) but a good many did appreciate the need to find the value for *v* and often went on to obtain an answer rounding to 4.3 as well.

Question 5

The use of interpolation, which was expected in part (a), is improving and many candidates made a reasonable attempt. Some failed to use correct class limits and had a width of 9 rather than 10 but the correct answer was often seen.
Despite the values for $\sum t$ and $\sum t^2$ being given in the question a number of candidates chose to estimate these values from the table and subsequently obtained an incorrect estimate for the mean but the mark scheme did permit them to obtain the remaining marks. The question did not ask for estimates of the mean and standard deviation and so using the mid-points was inappropriate in this question. The usual crop of errors arose when calculating the standard deviation with many failing to divide the 69378 by 32. Candidates can usually calculate S_{xx} correctly using the given formula and they may find it helpful to simply remember that standard deviation is simply $\sqrt{\frac{S_{xx}}{n}}$ but few seem to use this approach.

In part (c) most compared the mean and median and correctly concluded that the skewness was positive. Some used a formula which, apart from the extra work, was fine but those who used a (median – mean) formula rarely gave the correct conclusion. Some candidates went to great lengths to calculate the quartiles and conducted a quartile test for skewness which, if correct, led to a conclusion of negative skewness. This was allowed but was clearly not the intended approach for 2 marks.

Question 6

Most candidates showed us clearly that they were using the sum of the probabilities to reach $10k = 1$ from which they showed that $k = 0.1$ and the calculation of E(X) was usually correct too. In part (c) some confusion between E(X^2) and Var(X) and also over what to square (some choosing the probabilities rather than the values of x) caused difficulties and a few simply squared their answer to part (b). Part (d) was a fairly standard request and most knew that Var(2 – 5X) = 25Var(X) and were sometimes able to recover from errors made in part (c) but then many solutions ground to a halt.

Many candidates did not recognise the 3 cases required in part (e) and were therefore unable to emulate this approach in part (f) but it was encouraging to see that some of these candidates did realise what was required for part (g) and often gained both marks here.

Question 7

Many candidates were able to complete the tree diagram correctly but common errors were to have probabilities of 5/9 and 4/9 on the top two and bottom two branches. Part (b) was often answered correctly and even those with incorrect tree diagrams could achieve 2 or even all 3 marks here. Few candidates explained which four probabilities they were using to answer part (c) and the examiners were often left trying to deduce this from their tree diagram. Simply writing P(*RRR*) + P(*RYR*) + P(*YRR*) + P(*YYR*) would have earned them the first mark and made their solution much clearer. Some students failed to appreciate that a sum of 4 products of 3 probabilities from their tree diagram was required and a popular "fiddle" was to calculate $\frac{1}{4}\left(\frac{6}{9}+\frac{5}{9}+\frac{5}{9}+\frac{4}{9}\right)$ using the probabilities from the third branches and magically dividing by 4 to reach the printed answer. Part (d) was answered poorly with most attempts assuming that *A* and *B* were independent events (they were but this was never justified). Part (e) though was usually answered well with candidates clearly using the addition rule and the given answers. In the final part many candidates failed to identify the conditional probability and those who did often did not clearly state what their ratio of probabilities represented: once again a statement such as $("P"(RRR))/("P"(RRR)+"P"(YYY))$ with some suitable probabilities attempted would have secured the first mark.

Question 8

There were some good responses to this question and even some of those who had struggled with parts of questions 6 and 7 were able to pick up a good score here.

Part (a) was answered well and it was good to see diagrams being used to assist the candidates. Part (b) still causes problems for many candidates. They can usually standardise but then far too often equate this expression to 0.01 or 0.99. Those who did use a *z* value often used the "small" table and found the 2.3263 value but the minus sign was often missed and the final answer was therefore incorrect. Most seemed to try part (c) and the standardising was usually correct and suitable *z* values were often seen (use of 2.32 or 2.33 and 1.28 were acceptable here) but, when suitable equations were formed, a minus sign was often missing from the second equation. Solving their two linear equations was carried out quite well but full marks were only secured by those who worked carefully and accurately throughout this part.

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 publications@linneydirect.com

Order Code UA026663 January 2011

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

Edexcel Limited. Registered in England and Wales no.4496750
Registered Office: One90 High Holborn, London, WC1V 7BH