edexcel

Examiners' Report/
Principal Examiner Feedback

June 2011

GCE Statistics S1 (6683) Paper 1

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, 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/

June 2011
Publications Code UA028836

Statistics Unit S1
Specification 6683

Introduction

Examiners were pleased to see nearly all candidates tackling the standard calculations in questions 1, 7 and 8 confidently and accurately and the questions on the normal distribution, often a poorly answered topic, showed some improvement although candidates still struggle to present their work on this topic clearly.

Report on individual questions

Question 1

Part (a) was answered well with only a small minority using $4305 - \left(\frac{181}{8}\right)^2$.

Substitution into the formula for r was carried out successfully but a number of candidates gave their final answer to only 2 significant figures instead of the standard 3 significant figures we look for on S1. Most candidates now realised that the instruction "interpret" requires a contextualised comment but there were a number of nonsensical comments such as "temperature increases as sea level decreases" which gained no credit. Most candidates knew that coding had no effect on the correlation coefficient and picked up the mark for part (e) but very few scored the mark for part (d) with the commonest error being to divide by 1000. It appears that the effect of coding is being remembered as a fact rather than being deduced from an understanding of the structure of the formula.

Question 2

Most candidates knew how to standardise and there were very few dividing by 5^2 instead of 5. The usual problem arose with confusion of probabilities and z values and many simply equated their standardised expression to 0.9192. Those who did use 1.4 though invariably solved their equation successfully to reach 16.

Part (b) was a 1 mark question but some candidates wrote several lines with various degrees of success. Common errors were to give 0.9192 or 1 – 0.9192.

Question 3

The quality of responses to this question varied considerably. Most candidates stated that $a = 0.1$ and deduced that $a + b + c = 0.7$ but many were unable to proceed sensibly beyond this point. Some assumed a, b and c were all equal, others had $b = 0.2$ and $c = 0.4$ and some tried to consider E(Y) which led to some complicated and unhelpful equations. Finding d produced a number of errors: some confused P($Y = y$) with F(y) and gave answers of 0.3 or 0.7 whilst others used F(1) + F(2) + P(Y = 3) to obtain 0.9.

Part (b) proved challenging for many. Some solved the inequality to arrive at P($Y \geq 2$) and others created lists of all possible values for $3Y + 2$ before deciding which cases to include. Obtaining the answer from this point should have been straightforward but some simply stopped at this point or then found P(Y = 1) + P(Y = 2). Some candidates were confusing E($3Y + 2$) and P($3Y + 2$).

Question 4

Part (a) was tackled well by the majority of candidates although some were still unsure whether to give the answer as 0.7580 or 1 – 0.7580. There were a surprising number who mis-read the 16.12 value as 16.2 and lost the final accuracy mark.

In part (b) a number of candidates are still not using the table of percentage points to obtain the more accurate z value of 0.5244 but, as usual this only incurred a one mark penalty. Those who standardised and equated to a z value often failed to consider the sign of this value carefully and obtained the commonly seen answer of 16.96.

Question 5

Part (a) was, unsurprisingly, answered very well and most candidates were able to use the mid-points to find the mean in part (c). The calculation of standard deviation is still causing problems though: an incorrect formula, premature rounding and forgetting the square root were the typical errors. The technique of interpolation required in part (b) is now better understood but incorrect class widths still caused problems here. A common response to part (d) was to simply state that the data was continuous rather than using the calculations they have just performed in parts (b) and (c).

Part (e) was usually answered well with most using $Q_3 - Q_2 > Q_2 - Q_1$ but a small minority gave the wrong conclusion despite giving the correct comparison.

Question 6

It was encouraging to see many candidates drawing a correct Venn diagram for the start of this question but the wording "at least one" in part (a) was often misinterpreted as the intersection and some gave the probability of "only *K*" in part (b).

There were many correct answers seen in part (c) although many candidates simply wrote down the correct probability ratio without attempting an expression in symbols. Inevitably some wrote $P(J \mid K)$ but the identification and calculation of conditional probabilities seems to be improving.

In part (d) most candidates opted for the $P(K \cap J) = P(K) \times P(J)$ test whilst others attempted other incorrect comparisons such as $P(K \cap J) = 0$ or even $P(K \cap J) = P(K) + P(J)$.
Some merely quoted a standard formula in terms of *A* and *B* which, without some identification of what their *A* and *B* represented gained no credit. Only the very best realised that they had just calculated $P(K \mid J)$ and $P(K)$ and since these two were not equal the two events were not independent.

Part (e) discriminated very well with good candidates engaging with the context and showing they clearly understood the concept of independence.

Question 7

Parts (a) and (b) were answered well by the majority of candidates. Only a small minority used $\frac{S_{fh}}{S_{hh}}$ for *b* and there were few cases of the incorrect sign being used for finding *a*. Premature rounding and a failure to write their final answer in terms of *f* and *h* to 3significant figures meant a number lost the final accuracy mark. Part (c) was answered very well but in part (d) the candidates comments were often a little confused. The question was looking for a comment that the value of the independent variable *f* was within the range of the data and therefore the estimate should be reliable. A number of candidates seemed to focus their comments on the value of the dependent variable *h* and others were just a little vague referring to "it" was within the range of the data.

There were many clear and thorough answers to part (e) that showed the candidates had a good understanding of the limitations of the equation they had calculated (namely that it was based on data for children not adults). Questions of this type are simply looking for the candidates to engage with the context and give a sensible comment.

Question 8

The first 4 parts of this question were usually answered very well by most candidates. Very few confused $E(S^2)$ with $[E(S)]^2$ and the mean was usually squared when calculating Var(*S*). Part (e) was attempted by a good number of candidates and many realised that they needed to combine 0.2 and 0.2. Some gave 0.2 + 0.2 = 0.4 and others $0.2 \times 0.2 = 0.4$ but there were a number of correct solutions seen. Those who attempted part (f) often identified some of the cases such as $P(4 \cap 4 \cap 2) = 0.01$ and occasionally the 3 arrangements of 4, 4 and 2 but then they often missed $P(4 \cap 4 \cap 4)$.

Part (g) proved to be a challenging final part to the paper. Some tried listing all the possible cases for Jess to win. Their listing was usually extremely inefficient and they invariably missed one or more cases. Only a handful of candidates seemed to realise that only two probabilities needed to be considered namely $P(\bar{5} \cap 5 \cap 5)$ and $P(5 \cap \bar{5} \cap 5)$ and the probabilities could then be found quite easily.

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 UA028836 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