edexcel

Examiners' Report/
Principal Examiner Feedback

January 2012

GCE Statistics S1 (6683) Paper 1

Edexcel and BTEC Qualifications
Edexcel and BTEC qualifications come from Pearson, the world's leading learning company. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers. For further information, please call our GCE line on 0844 576 0025, our GCSE team on 0844 576 0027, or visit our qualifications website at www.edexcel.com. For information about our BTEC qualifications, please call 0844 576 0026, or visit our website at www.btec.co.uk.

If you have any subject specific questions about this specification 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/

Pearson: helping people progress, everywhere
Our aim is to help everyone progress in their lives through education. We believe in every kind of learning, for all kinds of people, wherever they are in the world. We've been involved in education for over 150 years, and by working across 70 countries, in 100 languages, we have built an international reputation for raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

January 2012
Publications Code UA030898

Introduction

The questions on the whole were well answered with many fully correct solutions. Weaker candidates found the paper very accessible and standard methods were well known and applied accurately. The paper discriminated effectively at the higher grades, especially question 2 and question 7.

Report on individual questions

Question 1

Part (a) was generally well answered, with most candidates getting the correct values. Only a few stated 7 and 1 (frequency density) and a few quoted other figures. A simple check that the total number of motorists should add to 100 would have told candidates if their calculations were incorrect. Some could not cope with the different interval widths and/or intervals starting and ending in ".5". This was more evident in part (b) where many struggled. The most common error was simply adding 21 and 45 not realising that 13.5 took them into the next group. A large number of candidates did realise that they needed to add three different frequencies together, although some found it difficult to get the correct fraction of the third class. Some ignored the third class or used its whole frequency.

Question 2

Despite the question using R and S in part (a) and *A* and *B* for the rest of the question, candidates assumed A and B were mutually exclusive and made no use of independence. In part (a) candidates were let down by their inability to express in unambiguous English "mutually exclusive". A number of candidates just restated the question, writing that it meant the probability of the intersection was 0 rather than describing the relationship between R and S.

Part (b) was not as well done as it ought to have been by the majority of candidates. Many didn't realise that the letters R and S were replaced from part (a) by A and B and so confused independence with mutually exclusive. Many candidates did write the full formula and substituted at least one probability correctly, although far fewer candidates realised the "independent" statement in the question meant that $\mathbf{P}(A \cap B)$ could be replaced with $\mathbf{P}(A) \times \mathbf{P}(B)$ Of those who successfully used the Addition Rule and independence, it was very disappointing to see some who could not handle the resulting linear equation because it had fractions in it. Those who did not start by quoting a formula and assumed exclusivity scored no marks.

Part (c) was answered well, with either a correct answer (even if part (b) incorrect) or a correct follow through.

In part (d) most knew that they had to use conditional probability, with only a few dividing by $\mathbf{P}(B')$ by mistake. The ability to find $\mathbf{P}(B' \cap A)$ for the numerator from previous working was often lacking and very few candidates used the fact that A and B were independent to simply state $\mathbf{P}(B' \mid A) = \mathbf{P}(B')$.

Question 3

Part (a) was answered well with a large majority setting out the solution as expected. A small number tried to verify the value, but most only did the substitution and did not say that it showed $k = 3$, thus losing the final accuracy mark.

Part (b) was poorly answered with a large number finding P(3) instead. A small number gave the answer as an inexact decimal instead of a fraction.

Part (c) and part (d) were both well answered with complete methods shown. Only a few candidates confused $[\mathrm{E}(X)]^2$ with $\mathrm{E}(X^2)$. In part (e) some of those candidates who got $\mathrm{E}(X^2)$ wrong still got $\mathrm{Var}(X)$ right here, as they did not realise the link and started again. Most realised that they needed to find $\mathrm{Var}(X)$ but many did not know the link with $\mathrm{Var}(7X-5)$. Some candidates worked out $7\mathrm{Var}(X)-5$, some $7\mathrm{Var}(X)$ and others $52\mathrm{Var}(X)$. The result for $\mathrm{Var}(7X-5)$ was often not awarded the final accuracy mark as some candidates had used rounded answers in their working.

Question 4

The first three parts were generally completed with confidence and fluency. Only a few candidates found incorrect values for the median (as they did $\frac{45}{2}$ and then looked for the 22.5th value) or the quartiles. The standard deviation was particularly well answered, but the usual errors of not dividing sum of x^2 by 45 or forgetting the square root were the most common.

In part (d) most candidates knew a rule to apply but a few got muddled with what exactly it was, using $Q_3 - Q_1$ or $Q_1 - Q_2$ or incorrectly for example. Having managed to apply the rule, some then got muddled in their reasoning, stating it showed positive skew or put the inequality signs the wrong way around which led to an incorrect conclusion. Those who simply used mean < median tended to fare better.

Part (e) was poorly answered, with some candidates not even attempting an answer. Perhaps they did not realise that this question considered a 'new' data set meaning it had little to do with previous work, with many performing calculations on the sample of 45 rather than considering the population. There were many scripts containing long and elaborate wrong calculations. The scaled mean was done better than the scaled standard deviation. On finding the new mean many candidates correctly subtracted 5 but multiplied their answer by 0.1 and did not subtract. A large number of those who attempted decoding applied the same rule to both the mean and standard deviation or stated incorrectly that the standard deviation was not affected by coding. Only a few candidates had intrinsic understanding that a measure of spread was only affected by the 'multiplier'. Those candidates who knew what to do produced short, efficient solutions.

Question 5

This question proved to be a good discriminator for the highest achieving candidates.
Part (a) and part (b) were well completed by most candidates, the biggest problem being giving r as -0.91 despite being told to give the answer to 3 significant figures.

Part (c) was generally very well answered but early rounding or lack of understanding of the difference between decimal places and significant figures led to the final equation not being stated accurately. Some candidates clearly did not know what an explanatory variable was in part (d), preferring to ignore the question and state what the dependent and independent variables were instead. A significant minority did not know what a variable was and suggested a or b.

In part (e) candidates who had the right equation had no problem getting the first answer, but a number then went on to find the answer when $t = 4$, rather than finding the change over four years. Some candidates seemed to think that they were just required to say whether the weight increased or decreased rather than find an amount, whilst others wrote a 'decrease of minus 0.1'. A few candidates believed that the coin could have increased in weight.

In part (f) some candidates just stated decrease or increase with no reason given. A number responded that it would increase and clearly understood that that the correlation would be stronger, but ignored that the result would make r closer to -1. A number of candidates said that the correlation was not affected by outliers and some thought that the removal of the fake coin constituted coding and so would have 'no effect on the product moment correlation coefficient'.

Question 6

Construction of the Venn diagram was nearly always correct. Occasional errors were mainly the omission of the box and failure to subtract frequencies accurately. Unfortunately, several candidates left the region for $R \cap S \cap C$ so small that it was extremely difficult to decipher the number written there.

In part (b) there were relatively few incorrect solutions. Occasionally an incorrect subtraction from 100 to find $n(R' \cap S' \cap C')$ was seen.

Part (c) and part (d) were very well answered by the majority of candidates. However, in part (d), $\frac{30}{100}$ was not an uncommon response, with the central frequency of 25 being omitted. This stems from a failure to understand the phrase "at least" in the question. Conditional probability in part (e) continues to be a problem for many candidates. Perhaps greater emphasis on the restricted sample space would produce better and quicker rewards.

Question 7

It was pleasing to see fewer blank pages than in the past although full marks were rarely gained for this question. Part (a) was answered well, if it was incorrect it was usually because candidates standardised with 25 as the standard deviation instead of 5 or did not subtract their probability from 1.

Part (b) was less well done and very few drew a diagram which helped with the areas and probabilities. Those who knew what to do often forgot to use the percentage points table and it was rare to see $z = 0.5244$ used. Some candidates used probabilities instead of z values, and some used the z value of 0.8416, although sign errors were few.

Part (c) was challenging for a large number of candidates and was not attempted if they had struggled with the earlier parts. The value of 0.16 was often seen in the scripts of those who tried it, ignoring the different ways of selection. Some candidates tried using the normal distribution to solve this part, or used values from part (b) such as 0.5793 in their work with lots of elaborate wrong calculations. Those who drew a tree diagram usually scored full marks.

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 UA030898 January 2012

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

Ofqual

Llywodraeth Cynulliad Cymru
Welsh Assembly Government

Rewarding Learning