Examiners' Report/
Principal Examiner Feedback

Summer 2012

GCE Statistics S1 (6683) Paper 01

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 visit our qualifications websites at www.edexcel.com or www.btec.co.uk for our BTEC qualifications.

Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

If you have any subject specific questions about this specification that require the help of a subject specialist, you can speak directly to the subject team at Pearson.

Their contact details can be found on this link:

www.edexcel.com/teachingservices.

You can also use our online Ask the Expert service at www.edexcel.com/ask. You will need an Edexcel username and password to access this service.

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 our commitment to high standards and raising achievement through innovation in education. Find out more about how we can help you and your students at: www.pearson.com/uk

Summer 2012
Publications Code UA033136

Introduction

The paper was generally accessible to all the candidates but question 5 proved challenging and over a quarter of the candidates could not get started. A small number of weaker candidates occasionally used negative probabilities or a correlation coefficient greater than 1 and the examiners would hope that they realised these answers were incorrect and showed some attempt to trace their mistake. The question on the normal was answered fairly successfully although candidates still struggle to present their work on this topic clearly. Some candidates still persist in giving final answers to 2sf rather than the standard answers which round to 3sf that we always expect in S1. This is a particular problem with correlation and regression questions.

Comments on individual questions

Question 1

This proved an accessible opening question to the paper. Part (a) was a "show that" and some candidates failed to show sufficient steps. There were two stages required: firstly the probabilities needed evaluating from the given probability function and most managed this successfully. Secondly there should be an explicit application of $\sum \mathrm{P}(X = x) = 1$ and some candidates failed to show this step clearly.

Most knew how to find E(X) in part (b) but some simply added their probabilities and a small minority divided their answer by 4.

Part (c) was another "show that" but most knew what to include here and full marks were often awarded. The use of notation was poor with a large number writing $\left(-1^2\right)$ when they actually meant $\left(-1\right)^2$. This was not penalised here but more attention to detail will be required as they progress to more advanced mathematical units.

For the final part most candidates now know the effect of coding on the variance and most realised that they needed first to find Var(X) and then multiply by $\left(-3\right)^2$ and many correct solutions were seen.

Question 2

This question was on familiar territory and was answered well. Most candidates could carry out the calculations in part (a) successfully although a small minority seemed unfamiliar with the formulae in the formula booklet and we had some using 23070 on the numerator of r. Some lost the accuracy mark for rounding to 0.82, or even 0.8, without first stating a more accurate value.

Part (b) was usually answered correctly but a significant minority launched into some complex decoding calculations or simply left it blank. The wording "write down" and the tariff of just 1 mark should indicate that complex calculations are not required.

In part (c) most recognised that a positive correlation suggested support for the bank's claim and scored both marks. Some just stated that the correlation was strong (failing to appreciate the importance of it being positive) and a few gave sociological rather than statistical reasons.

Question 3

Everyone could get started on this question but part (b) and drawing the line in (d) proved challenging for all but the best candidates.

Part (a) was answered well with only minor accuracy errors from some.

Few answered part (b) correctly with most trying to explain the purpose of linear regression rather than justifying its use in this instance.

Part (c) was answered very well and most knew how to use their answers in part (d) to find the gradient and intercept. A failure to work with sufficient figures on their calculators meant that many gave *a* as 0.738 which was acceptable for the first mark but not the 3sf, or better, accuracy we were looking for in the final answer. Very few gave their line as an equation of *y* on *x*.

In part (e) most plotted the mean point correctly but drawing the regression line accurately defeated many. A common error was to draw a line of best fit "by eye" that passed through the origin and a few weaker candidates simply joined up the dots.

In part (f) many correct answers were seen, mostly from substituting into the equation. Some used their graph (showing clearly on the diagram the line from p = 16 and from the regression line to the t axis) but this approach often lost the accuracy mark.

Question 4

In part (a) T and W or B and W were often stated as a mutually exclusive pair and a sensible reason given in words or symbols such as $\mathrm{P}(T \cap W) = 0$. Some candidates mixed up mutually exclusive and independent events and gave B and T as their pair with a reason that they could both happen together.

The test for independence in part (b) caused problems for some. The examiners were looking for a clear application of a suitable test which was usually $\mathrm{P}(B \cap T) \neq \mathrm{P}(B) \times \mathrm{P}(T)$, although some used a conditional probability successfully. We would also expect to see the required probabilities stated and this was a problem for some: a minority simply used integers and others gave $\mathrm{P}(B) = \frac{4}{25}$ or $\mathrm{P}(T) = \frac{3}{25}$. Those who carried out a correct test with correct probabilities usually gave a correct conclusion too and secured all 3 marks.

Part (c) was nearly always correct but a few gave $\frac{12}{25}$ as their answer to part (d). Most candidates now can identify a conditional probability and only a few attempted $\mathrm{P}(B|T)$ in part (e). It was quite possible to write down this final answer from the Venn diagram but many used the conditional probability formula from the formula booklet which if their $\mathrm{P}(B)$ was incorrect meant they could at least score the M1.

Question 5

This question was not answered very well. Many candidates either made a poor attempt at part (a) and then abandoned the question or just left it blank and moved on. Those who correctly formed a frequency table often scored well, whilst others who failed to complete part (a) struggled to make any headway with the remainder of the question.

In part (a) there needed to be some attempt to count squares and 22.5, 562.5 or 112.5 (small squares greater than or equal to 35 mph) were frequently seen. However many candidates did not appreciate the key idea that area is proportional to frequency and there was no attempt to combine this figure with the total frequency of 450. Those who did combine their figures and were able to come up with a correct relationship between area and number of cars (e.g. 1 large square represents 20 cars) were usually able to complete this part successfully although a few found the number of cars speeding above 30 mph instead of 35 mph as required. A few candidates stumbled upon 90 by dividing the 450 cars by the 5 bars of the histogram but this, of course, received no credit.

In part (b) most attempts tried to use mid-points but many struggled to find suitable frequencies and a few were unsure of the class widths (using 6 and 11). Some used the number of squares as frequencies but they rarely had a compatible denominator for their expression for the mean.

Most attempts at part (c) realised that interpolation was required but many promising solutions used 19.5 or 20.5 as class boundaries rather than 20.

Although they may not have had the correct values for the mean and median many had some values which they could use to answer part (d). A simple comparison of their values (e.g. mean greater than median) earned them the first mark and then an appropriate comment about the skewness (such as positive skew) the second. Some attempted to calculate the quartiles and invariably these were incorrect. Candidates should consider the amount of work involved in finding these values and compare it with the tariff for the question: 2 marks for a comment and a reason should not involve half a page of calculations. Other candidates tried to justify their comment from the shape of the histogram ignoring the calculations in (b) and (c).

In part (e) we required a choice of mean or median that was compatible with their conclusion in (d). Some candidates who had correctly concluded that the distribution was skewed in (d) still chose the mean, because it uses all the data, but there were many correct answers seen to this part.

Question 6

Most candidates now can standardise with confidence but there is still some confusion when using the tables for negative *z* values and a number of candidates are still not clear about the difference between a probability (area under the curve) and a *z* value (the point on the horizontal axis). The notation is still not handled well and statements such as "$\mathrm{P}\left(Z > \frac{150-162}{7.5}\right) = -1.6$" or "0.4 = 0.2533" are often seen as candidates stumble towards a correct answer.

In part (a) the majority gained the first 2 marks for standardising and arriving at a correct value for *z* but many lost the final mark for subtracting 0.9452 from 1. A clearly labelled diagram may have helped them avoid this error.
In part (b) some were confused by the reference to 60th percentile and simply found 60% of 162. Whilst many could standardise correctly equations such as $\frac{s-162}{7.5} = 0.6$ were quite common. Those who did realise that $\frac{s-162}{7.5}$ should be set equal to a *z* value often used 0.25 rather than the more accurate 0.2533 available from the table of percentage points.

Similar problems arose in part (c) and there was the added complication of obtaining compatible signs for a correct equation. Some candidates made double sign errors when solving $\frac{162-\mu}{9} = 1.2816$ to arrive at an answer of 174 and this lost the accuracy marks as their answer did not follow from their equation. The use of $\pm$ 1.28 was condoned here if the candidates had already lost a mark for using 0.25 rather than 0.2533 in part (b).

Question 7

This question was answered well by most candidates. There were a number of accuracy errors with the wrong positioning of the decimal point with
0.03×0.3 giving rise to 0.9, 0.09 or even 0.0009 and 0.047…was sometimes copied from a calculator as 0.47…
Part (a) proved a straightforward start to the question and most candidates completed a correct tree diagram.

Most completed part (b) correctly too but some found P(a split) and others found P(at least one of the defects) but many correct answers were seen here.

In part (c) the majority were able to multiply together the 3 required probabilities but a small minority just gave the answer as 0.97×0.98.

Numerous candidates found the correct products of three probabilities in part (d) but some just added 0.05 to their answer to part (b). Most candidates though found this question a very rewarding end to the paper.

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 UA033136 Summer 2012

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