edexcel

# Examiners' Report/
# Principal Examiner Feedback

## Summer 2016

Pearson Edexcel GCE
Statistics 1
(6683/01)

**Edexcel and BTEC Qualifications**

Edexcel and BTEC qualifications are awarded by Pearson, the UK's largest awarding body. 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. Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

**Pearson: helping people progress, everywhere**

Pearson aspires to be the world's leading learning company. 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 2016
Publications Code 6683_01_1606_ER

**Principal Examiner's Report**

**GCE Mathematics**

**Statistics 1 (6683/01)**

## Introduction

The question paper proved accessible to all the candidates but questions 2, 4, 5 and 6 provided plenty of discrimination for the top grades.

## Comments on individual questions

## Question 1

This proved to be a good opening question with nearly all students making some progress. Part (a) was answered successfully with most writing down a correct expression which was, of course, required for a "show that" question. Some forgot to write down the "= 5601" and lost the second mark.

In part (b), most chose the variable *w* but their reasons were not always precise enough. The question was phrased using the word "response" and so the examiners were looking for an explanation of this, rather than a statement such as "*w* responds to *d*". The key idea is that the number of wiggles depends on the distance of the food supply from the hive and this was all that was required. Some gave good descriptions of correlation but failed to mention the underlying causal relationship.

Part (c) was, as usual, answered very well but there were still a number of students only giving their answer to 2 significant figures and not the 3 significant figures that is standard on S1. More than 3 significant figures is fine, but answers given with fewer than 3 significant figures will lose a mark.

The calculations for part (d) were carried out confidently but a number lost the final accuracy mark by failing to carry enough figures in their calculations to achieve 3sf accuracy in the final equation and a few confused 3 significant figures with 3 decimal places. Most had a sufficiently accurate regression equation to obtain the mark in part (e)(i) and many gave a correct answer to (ii) too: some referring to interpolation and others stating clearly that the distance was in the range of the data. Common reasons for failing to score this mark were simply stating that "it" was in the range of the data without specifying that "it" is the independent variable we are interested in.

**Question 2**

Most students found the first four parts of this question accessible and appeared to have been well prepared for it. Almost everyone was able to use the sum of the probabilities condition to obtain a first equation in part (a) and there was often a correct expression for E($X$) seen too, although sometimes this was set equal to 1 not 0.4

Errors in solving the equations in part (b) sometimes led to negative values for the probabilities of $p$ and $q$ and it is disappointing to see students still using these impossible values in later stages of their working. One would hope that they would realise that an error had been made and then try to correct it.

The variance calculation in part (d) was usually correct with only a small number of students making slips, such as subtracting the mean instead of the mean squared, or dividing by 5.

Part (e) caused problems for a significant number of students. A surprising number did not try listing the probability distribution for $R$ and there were many cases of students trying to use the reciprocals of the probabilities or simply thinking that $\mathrm{E}\left(\frac{1}{X}\right)=\frac{1}{\mathrm{E}(X)}$.

Part (f) was only answered successfully by the more able students. Some students embarked on a correct approach but failed to correctly identify the cases where Sarah or Rebecca would win, a few forgot that there could be a draw and having found a correct answer to part (i) simply subtracted this from 1 to give their answer to part (ii). A number did not read the question carefully enough and thought that there were two independent values of $X$ being used: some then assumed that all 25 cases were equally likely but others battled through to obtain a correct response to this situation and a special case in the mark scheme enabled them to have some credit.

**Question 3**

There seemed to be a slightly improved understanding of the calculation of standard deviation this summer and many correct responses to part (a) were seen. The usual errors still occurred though: forgetting the square root, or failing to divide the 985.88 by 8 or simply confusing $S_{xx}$ with the variance.

The responses to part (b) confirmed that few students are aware of the simple relationship between $S_{yy}$ and standard deviation as many attempted to find $\sum y^2$ by "unpacking" the value for $\sigma_y$ with varying degrees of success. The calculation for $S_{xy}$ in part (c) was carried out successfully by most students and many completed part (d) correctly too, usually by first finding $S_{xx}$ rather than using their answer to part (a).

Responses to part (e) were very mixed. Some simply quoted a statement about coding not affecting correlation and therefore expected no change. Others realised that the new point was breaking the trend (both values above their means conflicting with the negative correlation) but were not precise enough about the effect on the value of the correlation coefficient. The examiners were looking for a statement about the value that told us that the correlation was weaker, i.e. that the value was closer to zero: $r$ "increases" or $r$ is "less negative" were not deemed to be satisfactory.

**Question 4**

Again it was disappointing to see some students working with negative probabilities but most could at least make a start here. Some floundered over the idea of *B* being a subset of *R* whilst also intersecting *D* and the final part proved quite challenging, enabling the more able students to shine.

Part (a) was usually answered correctly with 0.33 or 0.6 being the common incorrect answers. Many were successful in part (b) too but a common error arose from poor use of brackets: students arrived at $0.6(0.42 + t) = 0.27$ but then forgot to multiply the "$t$" by 0.6. Part (c) usually followed as most realised that $u + t = 0.25$.

The conditional probabilities in part (d) proved to be beyond some students and a common error in part (i) was forgetting to add the 0.33 to the 0.27 for the denominator and just giving an answer of 0.27/0.33 or $\frac{9}{11}$. Part (ii) was generally answered more successfully than part (i) but an incorrect denominator was often seen here too. Part (e) discriminated well with many simply trying $77\times0.45$ or $0.27\times40+0.15\times37$ and only the more able students using their answers from part (d) as required.

**Question 5**

The comment requirements in parts (e) and (f) made the latter stages of this question quite demanding but many were able to pick up marks in parts (b), (c) and (d). In part (a) most identified the width as 0.5 cm but, as usual, the height caused some difficulties. A number of students still seem to think that the height represents the frequency and some managed to make the statement that 4 $cm^2$ represents 8 babies but were still unable to use this to find the correct height.

The linear interpolation in part (b) was usually carried out very well but a few used the midpoints or made errors with class boundaries or used incorrect cumulative frequencies such as 10 or 27 instead of 9 and 26.

In part (c)(i), some students did not show how to use the midpoints to find $\sum fx$, simply using the value of 171.5; a "show that" question always requires a full explanation of the method used. More students now seem to know a suitable formula for standard deviation but some are still a little uncertain about whether or not a square root is required and some wasted time by calculating $\sum fx^2$ (not always successfully) rather than using the value given in the question.

Part (d) was a simple normal distribution probability calculation and the success rate amongst students with this standard application has improved considerably. A small minority are still uncertain about whether or not to subtract the value from the tables from 1 but there were many fully correct responses to this part.

In part (e), a number made a correct observation about the mean and median and possible skewness and could relate this to the suitability of using the normal distribution. Far fewer, though, were able to compare their answer to part (d) with the equivalent proportion from the table and so missed out on the other two marks. In part (f), most stated that the mean would not change and often gave a correct reason for this but fewer knew that the standard deviation would decrease and fewer still were able to give a convincing reason to support this.

**Question 6**

Overall, the standard of work on the normal distribution was much better compared to previous years. Most students now standardise correctly, although the notation is not always handled well, and the distinction between probabilities and *z*-values still causes problems for some students.

Part (a) was answered well by most students but fewer secured all 3 marks in part (b). Some used a
*z*-value of 0.84 rather than the more accurate value of 0.8416 from the percentage points table. The other problem was a misunderstanding about the concept of "high performance". To finish in the top 20% of runners, Nathaniel would be aiming for an answer <u>less</u> than the mean of 240 minutes. A number of students made a sign error and obtained an answer of 274 minutes and failed to realise that this must be incorrect.

Part (c) caused difficulties for many students. The best and simplest solutions were based on a diagram and appropriate use of symmetry. Others used their definitions of conditional probabilities but were unable to identify that their numerator of $\mathrm{P}([W < 30-\mu] \cap [W < \mu])$ simplifies to $\mathrm{P}(W < \mu - 30)$ and often tried multiplying two probabilities to obtain the numerator. A significant minority of students did not attempt this part but this appeared more to do with the difficulty of the question as opposed to an issue of time.

**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

Pearson Education Limited. Registered company number 872828
with its registered office at 80 Strand, London WC2R 0RL