www.mymathscloud.com

Pearson

Examiners' Report
Principal Examiner Feedback

Summer 2017

Pearson Edexcel GCE Mathematics

Statistics S1 (6683)

edexcel

www.mymathscloud.com
My Maths Cloud

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 2017
Publications Code 6683_01_1706_ER
All the material in this publication is copyright
© Pearson Education Ltd 2017

www.mymathscloud.com
My Maths Cloud

Statistics 1 (6683) – Principal Examiner's report

General introduction

The question paper proved accessible to all the candidates but questions 3, 5 and 6 provided plenty of discrimination for the top grades.

Question 1

This proved to be a good opening question with nearly all the students making some progress. Part (a) was answered very well but in part (b), few could work out the effect of the coding on S_{ss} although more were successful with S_{st}. Some reminded us that coding had no effect on the correlation coefficient and assumed that this must also apply to these values as well, whereas others embarked on long, and usually totally incorrect, calculations having missed the significance of the instruction to "write down…".

In part (c) most tried using a correct formula and many, quite sensibly, used the variables w and t but often their values of S_{ss} and S_{st} gave an answer of $|r| > 1$ and this also meant they lost then chance of scoring in part (d). It is disappointing that many students still don't seem to realise that $|r|$ cannot be greater than 1 and use this fact to try and fix their previous errors. Those who had a valid answer for r were usually able to give a suitable comment in part (d) though some failed to emphasise that it was the magnitude of the correlation, rather than the sign, that was significant. Part (e) was often fully correct but common errors were to use $b = \frac{-49}{50}$ presumably caused by a failure to identify the response and explanatory variables correctly. The usual loss of a mark for accuracy occurred where students prematurely rounded their gradient. A few students still don't appreciate that values of a and b given as fractions are inappropriate but it was encouraging to see very few cases of students giving their final equation in terms of x and y rather than t and w as required. There were many correct responses to part (f), but some did not spot the simple link to part (e) and attempted to calculate a new regression equation usually with little success. The responses to part (g) were disappointing with many simply substituting t = 1 into their equation rather than giving a suitable interpretation of the gradient. Others failed to give a full enough interpretation and simply said "a decrease" whilst other were confused over the units and were happy to quote a loss of 6.5p per week.

www.mymathscloud.com
My Maths Cloud

Question 2

Most could find the correct width in part (a) but the height calculation is still causing problems. The common error was to compare frequencies rather than frequency densities and an incorrect answer of 3 was fairly common. Part (b) was answered well by the majority of students with very few cases of students using (n + 1) and most choosing to work upwards from 45 rather than down from 50. The interpolation technique is understood but there were still a number of errors made in determining the appropriate class boundaries with 44.5 being used instead of 45 quite often.

As expected, part (c) was answered very well by the majority of students but there were still some who simply added the 5 midpoints and divided by 5 which is disappointing for students at this level. Many students didn't seem to be familiar with the convention that answers involving money should be given to 2 decimal places and some lost a mark for simply writing down £49.1 without any correct working being seen. There were many fully correct answers to part (d) as well, but some forgot to use the square root and a number lost accuracy by using 49.1 or the mean rather than the full expression.

In part (e), all that was required was a simple comparison of Q_2 with $\bar{x}$ but a number used a formula for skewness and that was, of course fine. The expected conclusion was that there was very little skewness since $Q_2 \approx \bar{x}$ but many concluded that since $\bar{x} > Q_2$ there was positive skew and this was allowed. The idea of using a normal distribution to <u>model</u> these data being an acceptable approach because there was only a little skewness was not appreciated by many students who simply trotted out a standard response that a normal was not a suitable model because the data was skew. Students should be encouraged to consider the magnitude of the skewness and realise that, with such a small amount of skew, to model the situation with a normal distribution may be a reasonable thing to do. Some students completely missed the need to reference part (e) in their answer to part (f) and simply compared their mean and standard deviation to Rika's values of £50 and £10. The final part (g) defeated many students despite being a fairly standard normal distribution question, presumably because it was not part of <u>the</u> question on the normal distribution. The three common errors were: standardising with their mean and standard deviation rather than the given values of 50 and 10, equating $\frac{c-50}{10}$ to a probability rather than a z-value or using 0.84 rather than 0.8416 from the percentage points tables.

www.mymathscloud.com
My Maths Cloud

Question 3

Although over a quarter of the students scored full marks, clearly reasoned and fully correct answers to this question were relatively rare. Most started correctly in part (a) and the correct values for p and q were often seen. Frequently there followed in parts (b) and (c) a string of equations linking r, s and t often accompanied by failed attempts to solve them correctly. The most common error though was to assume that A and B were independent which quickly gave an incorrect value for s and an easy route to find wrong answers for r and t. This also meant that any argument they used in (d) was flawed and marks were lost here too.

In part (d) those who had a correct (or nearly correct) Venn diagram were familiar with the two approaches to determining independence but by far the most popular was to compare $\text{P}(A)\times\text{P}(B)=0.5\times0.6=0.3$ with $\text{P}(A\cap B)=s=0.28$ and conclude that A and B were not independent. In the final part (e), most realised that a ratio of probabilities was required and many had a suitable denominator from $\text{P}(A) + \text{P}(C)$ or their $(r+s+p+q)$ but the numerator was often incorrect. A common error was simply to use $\text{P}(B)$ and many also used $\text{P}(B)\times\text{P}(A\cup C)$ and a few lost the final accuracy mark for simplifying $\frac{0.43}{0.75}$ to 0.57.

Question 4

This question clearly divided the students into two groups: those who understood the concept of the cumulative distribution function and in particular the difference between $\text{P}(X = x)$ and $\text{F}(x)$ and those who did not. Nearly 40 % gained full marks here but over 15% scored zero. Part (a) was often fully correct though a common error was to have $b = \frac{1}{6}$ and $c = \frac{1}{3}$ but these students often picked up a follow through mark for d and sometimes the mark in part (b) as well.

Part (b) proved quite challenging with 30% of students scoring all the marks in part (a) but then failing to get part (b). The problem seemed to be students not realising that $\text{P}(X^2 = 1)$ meant $\text{P}(X = 1$ <u>or</u> $X = -1)$ and common incorrect answers were $\frac{1}{16}$ from $[\text{P}(X=1)]^2=\left(\frac{1}{4}\right)^2$ or $\frac{1}{12}$ from $\text{P}(X=-1)\times\text{P}(X=1)$ rather than $\text{P}(X=-1)+\text{P}(X=1)$.

www.mymathscloud.com

Question 5

This was the most challenging question on the paper. Part (a) was a simple use of the normal distribution and most students scored full marks here. A sizeable group though standardised correctly but then confused the two tables and wrote down the z-value corresponding to a probability of 0.4 (i.e. 0.2533) and used this as their probability. There were also a number of cases of students miscopying the tables and 0.3466 instead of 0.3446 was seen quite frequently. The usual errors of standardising with 5^2 or $\sqrt{5}$ were very rare though a number were still unsure whether or not to subtract their tables figure from 1.

In part (b), few wrote down a conditional probability statement but launched straight into a ratio of probabilities. This was fine if the ratio was correct but, whilst most had $P(T > 15) = 0.7257$ on the denominator, those who thought the numerator was simply $P(T > 15) \times P(T > 20)$ would score no marks. Those who did have a correct numerator were often able to score all of the marks here. Part (c) proved too demanding for most students and only a few of those who eventually arrived at a correct answer did so from a clearly argued solution. Some students assumed that linear interpolation was required and made no headway. Very few started by considering a conditional probability but some did realise that they required the value d such that $P(T > d) = 0.5 \times 0.7257$ and some then realised they were trying to find d from $P(T < d) = 0.63715$. Those who got this far usually realised that they needed to use a z-value of 0.35, though some did not know how to deal with a probability in-between two values in the tables and simply took the midpoint, and were able to form a correct equation and arrive at the correct answer but under 10% scored full marks on this question.

www.mymathscloud.com

Question 6

Although the final part of this question was quite challenging most students were able to get started and most scored at least half marks. Parts (a) and (b) were straightforward and nearly all scored full marks here. A small minority confused E(X^2) with Var(X) and some forgot to square the E(X) term when finding the variance. Part (c) should have been straightforward but some students ignored the instruction to form a linear equation in p and simply substituted $p = \frac{1}{3}$ and $q = \frac{2}{3}$ into the formula for E(Y) but gained no marks. Some used $p + q = 1$ and started with two simultaneous equations which made the question a little more complicated than intended.

Part (d) was usually answered correctly but some were incorrectly using a random variable X here. Many were less successful in part (e) because they failed to present a clearly argued "show that" solution. Simply multiplying $\frac{1}{3} \times \frac{1}{4}$ was not sufficient. We required some indication that the event $S = 30$ was achieved when $X = 6$ and $Y = 5$ as well as the multiplication of $\frac{1}{3} \times \frac{1}{4}$. Many students made no attempt at the remainder of the question but those who did could usually make some progress. Those who drew up a distribution table in part (f) usually scored at least the first mark but there were often errors in P($S = 4$) or P($S = 25$) and sometimes extra values crept in too such as $S = 0$. Those with a correct table in (f) were usually able to find E(S) correctly although a small number failed to give an exact answer which we always expect for work on discrete random variables. Part (h) was the most difficult part of the question. Many students ignored the fact that Sam and Charlotte play the game a large number of times and compared individual outcomes or the probability of achieving a higher score rather than comparing the expected scores for each. Some did not realise that Charlotte's expected score had already been calculated in part (b) and found E(X^2) again, sometimes incorrectly this time, and others thought they should compare E(S) with $[\text{E}(X)]^2$ but a very competent 10% of the students gave a fully correct solution to this question.

www.mymathscloud.com

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