edexcel

Examiners' Report/
Principal Examiner Feedback

Summer 2013

GCE Statistics S1 (6683)
Paper 01

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications come from 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 2013
Publications Code UA036992

Statistics S1 (6683)

Introduction

The paper was accessible to most candidates and questions 1 (on correlation and regression) and 2 (on box and whisker plots) were answered very well but the normal distribution (question 6) still proved to be quite discriminating. Sometimes answers to questions were wrongly labelled and this can make it difficult for the examiners to find the appropriate working; candidates should heed the advice on the front of the papers. The usual problems of premature rounding arose again and candidates should be aiming to keep sufficient figures from preliminary working in order to present their final answers to 3 sf or better.

Report on individual questions

Question 1

Part (a) was, as usual, answered very well but a number of candidates lost the final A1 because they did not round their answers to 3sf or, more worryingly, they thought that S_{hh}= 149 to 3sf. Most knew how to calculate *r* in part (b) too but few gave a full answer to part (c). Many stated that there was negative correlation (although some thought this meant that the use of a regression equation was <u>not</u> suitable) but few stated clearly that the use of a regression equation was suitable because there was <u>strong</u> correlation. Some simply said that "the points were close to a straight line" but there was no scatter diagram to support this and without a clear statement that the strong correlation suggests this the examiners could not award the mark.

Most candidates (even those who felt that a regression equation was not appropriate!) could carry out the calculations in part (d) although a sizeable minority used S_{tt} instead of S_{hh} which gave them a somewhat unrealistic gradient of – 60.3. Most found a correct gradient but often rounded their answer before calculating the intercept and the final mark was frequently lost. Full interpretations in part (e) were rare with candidates failing to mention the drop in temperature or the rise in height above sea level or give their value. The final part was answered quite well with most candidates substituting values of 500 and 1000 into their equation, only the better candidates realized that the answer was easily found from 500*b*. A number of candidates seemed perfectly content with a final answer of around 30 000 ^{o}C here (due to their incorrect gradient in part (d)) and lost the final mark. Candidates should be encouraged to try and engage with the context of the questions and this can help them both in interpreting their statistical calculations and assessing the reasonableness of their answers.

Question 2

There were many incorrect answers of 38 given to part (a) where candidates, presumably, thought the question wanted the value that 75% of candidates exceeded rather than the value exceeded by 75% of the candidates. They should be encouraged to read the question carefully.

Part (b) was answered very well and where they showed some calculations to determine the outliers they usually scored well on part (c) too. A surprisingly large number though failed to show any working for the outliers and their diagram showed no evidence of them either. There was still some uncertainty about where to end the whiskers (either at the outlier limits or at the next non-outlier value are acceptable) and there was some evidence that candidates had been using previous mark schemes as "model solutions" and were drawing both sets of whiskers which is, of course, incorrect. There were many good answers to part (d) but some candidates failed to secure marks as they did not use the appropriate terms. The use of "mean" or "average" or "spread" are not sufficiently clear: we require use of the correct terms such as "median", "range" or "Inter Quartile Range" .

Question 3

Part (a) was answered very well but in part (b) a number of candidates failed to spot or consider the complement (giving an answer of 0.01) and others confused the 200 with 100 and gave an answer of 0.98. The conditional probability in part (c) was answered quite well but a few had P(W) on their denominator and some assumed independence when calculating the numerator and used $\mathrm{P}(C)\times\mathrm{P}(W) = \mathrm{P}(C \cap W)$. Part (d) was a little different from the usual Venn diagram and candidates had to consider carefully how to represent the 4 events. Three overlapping circles or 3 separate circles with no indication of set B was quite common and those who did have a correct shape sometimes struggled to place the frequencies or probabilities. Those using frequencies were usually more successful as the probabilities were not always out of 200. A Venn diagram such as this should, of course, always have a box defining the universal set [and ideally a 0 for the region $\left(F \cup C \cup H \cup B\right)'$] and a few candidates missed this out. Despite their difficulties with the structure of the diagram for part (d) many candidates were able to interpret the table correctly and score the marks in part (e).

Question 4

A small number of candidates still failed to calculate the mean correctly in part (a). For some this was due to errors with the midpoints but the more extreme errors involved dividing by 6 rather than 200 or using the class widths rather than the mid-points. The standard deviation formula still causes problems for some: forgetting the square root and failing to divide $\sum ft^2$ by 200 were common errors and some candidates used their rounded value of the mean and lost the final accuracy mark as their answer was not accurate to 3 significant figures. The calculation of the median in part (b) was answered well but applying the same principles to part (c) caused difficulties for some with many of those attempting the $(n + 1)$ approach using 50.5 instead of 50.25 and others using incorrect end points. In part (d) a few spotted that Q_3 was on the class boundary and gave the value of 25.5 but others encountered similar problems to those with Q_1 but most were able to find their interquartile range. Part (e) was not answered well with many mentioning "continuous data" or "extreme values" and only a few stating that their data was skewed. Most candidates scored some marks in part (f) but many failed to secure all the marks because they did not deal with all of the estimates; in particular the standard deviation was often omitted.

Question 5

In part (a) most attempted an equation in a and b based on E(X) and, apart from those who forgot to multiply 6 and 0.3, this was usually correct. Many though failed to write down a second equation based on the sum of the probabilities although this often didn't deter them from "solving" for a and b and often arriving at the correct values although justification for this choice was rarely seen. Those who did obtain 2 simultaneous equations could usually solve them and a variety of methods were used. Those who had correct values for a and b had few difficulties in establishing the result in (b) and part (c) was answered well too with only a small number trying 5 – 3 Var(X) or $5^2\text{Var}(X)$.
The cumulative distribution function is still not understood very well by many S1 candidates and correct answers to parts (d) and (e) were rare. A common error was to assume the table was a probability distribution, set the sum of the F(y) values equal to 1 and arrive at $k = \frac{7}{120}$. A number did write P(Y = 1) = 0.1 and occasionally P(Y = 2) = F(2) – F(1) was used too but few achieved fully correct answers. Some candidates continued this process to find P(Y = 3) with their value of $k = \frac{7}{120}$ and seemed unperturbed by the negative value this gave.

The final part (f) was usually answered well and it is pleasing to see candidates carrying on to the end of a question even when they had encountered difficulties in earlier parts. Some simply added 0.1 and 0.1 and a few insisted on multiplying a correct 0.01 by 2 but the correct answer was often seen.

Question 6

As usual the normal distribution posed serious problems for some candidates. There were two major reasons for lost marks. The first was a failure to use the table of percentage points of the normal distribution where appropriate and many candidates lost marks for using z values of 1.28 or 2.32 or even 2.33 rather than the 4dp values available. The second problem is a fundamental one of understanding where candidates confuse probabilities (areas) with z values (points on the horizontal axis).

In part (a) many could standardize correctly and often set their expression equal to a suitable z value but often there was a sign error and this led to an answer of 190 for the mean. Other candidates stated that the mean was 210 which was correct but this didn't follow from their equation and accuracy marks will not be awarded in such cases. Part (b) was more straightforward and provided a correct mean was found in part (a) full marks were usually obtained. Many left their answer as a probability rather than the percentage asked for in the question but this was condoned on this occasion. The final part proved quite challenging. Some drew a diagram but were unable to represent the information in the question in a useful way. Others tried subtracting two standardizations and ended up with $\frac{10}{\sigma} = z$ which was of little use to them. Those who realized that just using the value of 210 along with the mean of 205 and $z = 2.3263$ was all that was required usually formed a simple equation for σ and were able to solve it successfully.

Grade Boundaries

Grade boundaries for this, and all other papers, can be found on the website on this link:
http://www.edexcel.com/iwant to/Pages/grade-boundaries.aspx

Pearson Education Limited. Registered company number 872828
with its registered office at Edinburgh Gate, Harlow, Essex CM20 2JE