edexcel

Examiners' Report

Summer 2015

Pearson Edexcel GCE in Statistics S1 (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 2015
Publications Code UA042707

Mathematics Unit Statistics 1
Specification 6683/01

General Introduction

The paper was accessible to all the students and very few students were unable to make some headway in any of the questions. The normal distribution (question 6) is still a stumbling block for many and the parts of questions requiring some interpretation or comment (questions 1(f), 2(d), 4(a), (d) and (f)) continue to provide some discrimination. Basic calculations though, like those required in questions 2(b), 4(b) and (c), are now carried out very well by almost all the students.

Report on Individual Questions

Question 1

In part (a) a large number of students ignored the outlier and gave an answer of 30 instead of 39. Others failed to score the mark because they left their answer as 48 – 9 or they gave two answers: one with and one without the outlier. Part (b) on the other hand was answered very well with most students scoring the mark. Part (c) caused problems for some students who did not use a correct lower class boundary with 64.5 being a common error. Some of these students managed to correct this error in part (d) and most were successful here. The given answer tempted some to take the midpoint of the 60-65 interval and round up to get 63. We allow students to use (n + 1) instead of n when calculating the cumulative frequency to use in the interpolation. Most used 9.5 correctly in part (c) but in part (d) few choosing this approach used the correct value of 15.25. In part (e)(i) many students were able to find the correct limits of 93 and 45 and make a suitable comment about the outliers. A sizeable minority though used the interquartile range from part (b) instead of the correct value of 12 and lost these marks. The box and whisker plots were usually correct with only a few making careless errors such as plotting the top whisker at 85 instead of 84. The final part proved to be a good discriminator with many revealing a lack of understanding about how to interpret the statistics they had calculated. Most achieved two of the 3 marks for choosing the 70^o angle and giving a suitable reason based on the closeness of the median to this value or the smaller range or inter quartile range but few gave both reasons. Some surmised that a value in the range 60-65 would be more accurate based on this being the modal class, others referred to the absence of outliers or tried to use symmetry to justify their choice.

Question 2

Most realised what was required in part (a) and although some went “round the houses” to solve their two simultaneous equations the correct answers were usually obtained. Part (b) was a simple application of the formula but sadly a number of students gave their answer as – 0.75 rather than an answer rounded to 3 sf. Most answered part (c) correctly with only a minority re-calculating the coefficient. Part (d) was not answered very well with many not appreciating that the correlation they had established was between distance from the station and price per square metre. Many thought they were looking at negative correlation between price and distance from the station and understandably chose *J*. Some students though did calculate the price/m^2 for each house but then some still chose house *J* as they could not interpret the negative correlation correctly but there were plenty of fully correct answers showing a thorough understanding of the situation. A few students ignored all the statistics and based their choice on other reasons such as it is “noisy near the station” and therefore the house would be cheaper.

Question 3

Part (a) was usually answered very well but a significant number did not appreciate that those who studied all 3 subjects were also included in the numbers studying a pair of subjects. Most could score the mark in part (b), even if their Venn diagram was incorrect, but some muddled their denominators and gave an answer of $\frac{13}{58}$ rather than $\frac{13}{80}$. In part (c) most wrote down a fraction based on their Venn diagram: this was fine if their diagram and answer were correct but otherwise a correct expression was required before the examiners could award any marks. In part (d) the examiners were often able to follow through a suitable numerator based on the student's Venn diagram and a denominator from their answer to part (c) and those who had scored both marks in (c) often did so again in (d). Some students though are still not spotting the conditional probability and denominators of 80 were quite common here. Some answers to part (e) simply lacked sufficient detail or explanation to award the marks. The examiners expect to see the relevant probabilities required for the test clearly stated and a correct test used with a concluding comment; simply writing "$\frac{20}{80}\times\frac{28}{80}=\frac{7}{80}$ so independent" is not sufficient. The most popular test was based on comparing $\mathrm{P}(B)\times\mathrm{P}(C)$ with $\mathrm{P}(B\cap C)$, though some used a conditional probability, but $\mathrm{P}(B\cap C)=\frac{4}{80}$ was a common error. A small number of students confused "independence" with "mutually exclusive".

Question 4

There was plenty in this question for all of the students with parts (b), (c) and (e) providing a good source of marks for everyone but (a), (d) and (f) proving to be more challenging.

In part (a) the first mark was available for a sensible reason and many were prepared to have a go at giving an answer here. Better students gave 2 clear reasons for using statistical **models** whereas many started mentioning reasons for carrying out statistical **tests.** Most scored full marks in part (b) but a number mis-read 255 as 225 and lost accuracy and others lost time by calculating S_{xx}. The calculations required in part (c) were well rehearsed but a few did not use the negative value of *b* correctly and some failed to write down the final equation or left their values of *a* or *b* as fractions. There were many correct answers to part (d) but some gave an interpretation of the gradient not the intercept and others simply said that 28.1 was the "*y* intercept" without attempting to give any "contextualised" interpretation. Part (e) was straightforward and nearly everyone scored the M mark here. Part (f) was often answered well as students realised that because the model was based on data collected in winter it would be unreliable to use it for summer temperatures as these would be different or involve extrapolation. The word "discuss" prompted some to surmise what the effect of air-conditioning units would be in the summer, and this often led to contradictory statements, whilst others felt that they needed to calculate the product moment correlation coefficient and then deduced that the model was reliable.

Question 5

In part (a) many realised that a score of 15 came from getting 2 correct answers and one incorrect answer but they just gave the probability as $0.6\times0.6\times0.4$ omitting the multiplication by 3. Others thought that the probability of 0.432 came from 2×0.216 and scored zero. Most obtained 0.288 in part (b) though usually, though not exclusively, by using the fact that the sum of the probabilities equalled 1. Some students argued in circles, using the given value of 0.432 to find 0.288 and then using their derived value of 0.288 to "find" 0.432, they of course, scored no marks for part (a). Part (c) proved to be quite discriminating. Many could identify <u>some</u> of the required cases and 0.216×0.288 was often seen or sometimes 0.432^2 but it was less common to see all 3 cases included. Some attempted to consider the situation as a set of 6 questions but they rarely considered all 6C4=15 arrangements. There was some evidence that students were mis-interpreting "a total of 30 points in 2 rounds" to mean "30 points in each of 2 rounds" and giving their answer as simply 0.216^2 : they should be encouraged to read and interpret the questions carefully. The methods for part (d) and (e) were well known and many scored well here. A number failed to use brackets carefully and found $-15^2 \times 0.064$ rather than $(-15)^2 \times 0.064$ but there were fewer cases than sometimes of students forgetting to square the mean before subtracting or thinking that $\text{Var}(X) = \text{E}(X^2)$ in part (e). In the final part most chose to form the distribution of Y = the number of points in a bonus round and hence find E(Y) using the given formula. A common mistake was to have 0 instead of 10 and a few students gave an answer of 35 since this value had the highest probability. A handful of students spotted that $Y = \frac{5}{3}X + 10$ and were able to write down the answer quite simply but this was very rare.

Question 6

Students who had been taught to use a diagram alongside the normal distribution tables were often able to make good progress in part (a). In part (ii) the common error was to find 1 – 0.9713 = 0.0297 and some mis-read the tables in the first two parts with answers of 0.1335 being quite common for (i). A clear diagram was usually a great help in part (iii) and there were a good number of correct answers here but part (iv) caused some problems. Some tried to use the addition formula given in the formula booklet but thought that $\text{P}(A \cap C)$ was $\text{P}(A) \times \text{P}(C)$ and the incorrect 0.8845 was common. Those who were able to interpret $A \cup C$ in terms of the inequalities could simply write down P(Z < 1.5) and hence find the correct answer from the tables.

Part (b) proved to be a good discriminator but many students failed to make progress due to poor handling of the notation or an inability to interpret the conditional probability. Most realised that at some stage they would need to standardise and find P(X > 28) but this was often the only mark scored. Those who realised that the given probability statement reduced to $\frac{\text{P}(X > w)}{\text{P}(X > 28)} = 0.625$ were usually able to make significant progress, the difficulty for many was appreciating that $\text{P}([X > w] \cap [X > 28])$ = P(X > w). Once they had got as far as P(X > w) = 0.625×0.0808 or P(X < w) = 0.9495 they were usually able to standardise and set equal to 1.64 and solve for w. Some used 1.6449 instead of 1.64 but this only cost them 1 mark and a reasonable number of correct answers were seen.

The notation for the normal distribution is still not handled well by many students at this level and there was a clear difference between those who gave a correct and rigorous argument leading to 29.2 and those who stumbled their way through despite several “nonsense” statements to arrive at the correct answer. This time the mark scheme did not differentiate between them but this is an area that the examiners should aim to address in future.

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, United Kingdom