Examiners' Report

Summer 2014

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 2014
Publications Code UA040116

Mathematics Unit Statistics 1

Specification 6683/01

General Introduction

The paper proved accessible to most students. It was noted that a number of students seemed unaware of the instruction to work to an "appropriate degree of accuracy". If students are using a calculator they would be well advised to first write down all the values on their display before rounding to 3 significant figures or whatever other accuracy is required. An answer which rounds to 3sf is usually accepted and so if an error is made in rounding, but a more accurate version is seen, then this can be given the credit.

Comments on Individual Questions

Question 1

Most students had a thorough understanding of box-plots and skewness and there were many fully correct answers. In Q01(a) the correct values were usually given but $a = 40$ and $b = 73$ were common incorrect answers where students failed to read the "leaves" from right to left .

In Q01(b) most students calculated the upper limit of 91 for outliers correctly and there were few errors in reading the scale and plotting 25, Q_1, Q_2 and Q_3 correctly. The upper whisker did cause some problems. An upper whisker stopping at 91 (the upper limit for outliers) or stopping at 75 (the next non-outlier in the data set) and then an outlier clearly shown by a cross or * at 99 was all that was required. Common errors were to draw both whiskers or to plot the outlier at 91 or simply to extend the whisker to 99 and indicate the outlier there too.

In Q01(c) most comments about the skewness of Penville were correct but some did not realise that they needed to comment on Greenslax too. The reasons were usually based on a comparison of the differences in the quartiles and were often correct although a number still think that $Q_3 - Q_2 > Q_2 - Q_1$ means negative skewness. Some students elected to calculate the means to compare with the medians, which was accepted if the means were correct, but others tried an argument based on the mode failing to realise that the distributions were bimodal and therefore this approach was invalid.

Question 2

Some students found this question challenging due to a lack of understanding about coding and others due to poor algebraic manipulation. Most appreciated that coding affected the mean, although they often substituted 60.8 for x rather than y . Those who started by writing down $60.8 = 1.4x - 20$ often managed to reach the correct answer but a number could not solve this equation for x correctly. Those who did not write this down knew that they had to add 20 and divide by 1.4 but could not get the order of operations the correct way around. Many thought that coding did not affect the standard deviation and others simply repeated the calculation for the mean with 6.60 rather than 60.8. Some students were not clear about the difference between standard deviation and variance and included factors such as 1.4^2 or $\sqrt{1.4}$ A few students gave their final answer as 4.7 rather than 4.71 and lost the final mark.

Question 3

The calculations were generally carried out very well here but the final 3 parts, requiring the students to engage with the context, were answered less well.

Despite mentioning the expectation of 3sf accuracy a number of students lost a mark here as the only evaluated answer they gave was 0.96 rather than 0.962. In Q03(b) most gave a comment about strong correlation or commented that the value of r was close to 1 but some misinterpreted the question and explained why linear regression was a useful tool rather than justifying its use in this situation.

In Q03(c) a number gave an answer of 0.74 rather than 0.740 as required but this value was allowed for the final equation in Q03(d) which many obtained correctly. It was rare to see a student give their equation in terms of y and x rather than m and v.

Q03(e) was found to be challenging. Many confused the gradient with correlation and simply said that as the number of visitors increased so did the amount of money spent. Those who did get the idea of rate often failed to give the correct numerical values or they had visitors and money the wrong way around e.g. "for every £1000 spent the number of visitors increased by 740" . The better answers identified that b represented the amount of money spent per visitor and gave simple answers such as "each visitor spends £740".

The most common error in Q03(f) was for students to simply substitute 2 500 000 into their equation but those who realised that 2500 should be substituted usually arrived at the correct answer and often went on to affirm the reliability of their estimate by pointing out that 2500 was within the range of the given data in Q03(g). Stating "it is within the range of the data" in part Q03(g) did not secure the marks unless the students made it clear that their "it" referred to the number of visitors and not the amount of money spent.

Question 4

The tree diagram was answered very well with only occasional errors on the branches for broken or not broken biscuits e.g $\mathrm{P}(B \mid J) = \frac{2}{25}$ rather than $\frac{2}{100}$.

Q04(b) and Q04(c) were usually correct although there were a number of transcription errors such as 0.335 instead of 0.0335 in Q04(c).

Q04(d) was a slightly more challenging conditional probability and it caused difficulties for some students. Some misinterpreted it and simply found $\mathrm{P}(K' \cap B)$ whilst others found $\mathrm{P}(K \mid B)$ and the error of having a numerator of $\mathrm{P}(K') \times \mathrm{P}(B)$ rather than $\mathrm{P}(K' \cap B)$ was not uncommon.

Question 5

This was another accessible question with only Q05(b) being challenging to some.

In Q05(a) most chose to write down the sum of the probabilities in terms of k and then set this equal to 1 and deduce the value of k. Some chose a "verification" route but often failed to give the final statement "therefore $k = \frac{1}{8}$" and lost the final mark. Many students still do not recognise the cumulative distribution function $\mathrm{F}(x)$ and there were many blank or incorrect responses here. Some confused $\mathrm{F}(5)$ with $\mathrm{P}(X = 5)$ and gave the answer 0 and others gave $\frac{5}{18}$. Finding $\mathrm{E}(X)$ and $\mathrm{E}(X^2)$ caused few problems although some students ignored the instruction to give the exact values and rounded their decimal answers.

In Q05(e) most knew how to find $\mathrm{Var}(X)$, although a few forgot to square $\mathrm{E}(X)$, and many knew how to deal with the $\mathrm{Var}(3 - 4X)$ formula with only some students trying $3 - 4\mathrm{Var}(X)$.

Question 6

Many students are still unsure about calculating the heights of bars in histograms but the work on linear interpolation and even standard deviation seems to be improving.

Most were able to state the correct width of the bar but few used frequency densities correctly to find the height, some finding the frequency density of $\frac{1}{3}$ but then calculating $\frac{1}{3} \times 2.5$ rather than $\frac{2.5}{\frac{1}{3}}$. Some identified that $1.5\ \text{cm}^2$ represented 10 customers but were then unable to use this correctly to find the height. Q06(b) was answered well but some students had an incorrect class width because they did not realize that the lower class boundary was 70 not 69.5. Q06(c) the mean was usually correct, although a few weaker students divided 6460 by 6 (the number of groups), and there were fewer errors made in calculating the standard deviation: some forgot the square root and others forgot to divide $\sum fx^2$ by 85. A few students ignored these given values and recalculated $\sum fx$ and $\sum fx^2$ (often incorrectly!) and wasted valuable time. In Q06(d) most were able to use the given formula with their values and most gave a sensible comment based on their evaluation.

Question 7

Most students standardised correctly although the notation, and especially the distinction between probabilities and z values, was not handled well.

In Q07(a) a small minority are still unclear when and when not to subtract the value found in the tables from 1.

In Q07(b) many found $P(H > 180)$ but they did not appreciate that Q07(b) wanted a conditional probability and so just left their answer as $P(H > 180)$ instead of dividing this by their answer to Q07(a). Some of those who did identify the conditional probability could not interpret their numerator $P(H > 180 \cap H > 170)$ correctly and simply wrote $P(H > 180) \times P(H > 170)$.

Q07(c) was often not attempted but those who did could usually obtain a probability of 0.0528 for $P(H > h)$. Unfortunately a number of students then used a rounded probability of 0.05 and therefore a z value of 1.6449 rather than the correct value of 1.62. Standardising and forming a suitable equation for h were usually accomplished correctly by those who reached this stage and the correct answer was seen from a number of students.

Question 8

This question was found to be challenging. A clear Venn diagram would have helped some to get started.

Those who tried to draw a Venn diagram could quickly find $P(A)$ from $1 - 0.18 - 0.22$ to answer Q08(a) and Q08(b) would either follow from $P(A) + 0.22$ or $1 - 0.18$. After this the students needed a clear argument to answer Q08(c) that did not assume that A and B were independent. Some students set off in the right direction by quoting the conditional probability formula $P(A|B) = 0.6 = \frac{P(A \cap B)}{P(B)}$ but to make further progress they needed a second equation with $P(B)$ and $P(A \cap B)$ which they could obtain by using the addition formula and their answers from Q08(a) and Q08(b). The more able students were able to successfully solve these two equations to find $P(B)$ or $P(A \cap B)$. Some students noted that since $P(A) = 0.6$ then if $P(A|B) = 0.6$ as well then A and B are independent, thus answering Q08(d) first. They could then use the given information $0.22 = P(A' \cap B) = P(A') \times P(B) = [1 - 0.6] \times P(B)$ to obtain $P(B) = 0.55$ and then their Venn diagram to see that $P(A \cap B) = 0.33$. Those who obtained $P(B) = 0.55$ from a more conventional route nearly always chose to answer Q08(d) by checking that $P(A) \times P(B) = P(A \cap B)$ and giving a correct conclusion.

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 Edinburgh Gate, Harlow, Essex CM20 2JE