Examiners' Report

Summer 2015

Pearson Edexcel GCE in Statistics S2 (6684/01)

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 UA041212

Mathematics Unit Statistics S2

Specification 6684/01

General Introduction

On the whole this paper was well answered. Most questions on the paper seemed accessible to the students. Generally the work was quite well presented although students did not always show their methods clearly for example not writing down the distribution they are using nor writing down probabilities associated with it.

The paper proved accessible to the majority of students and there was little evidence of there not being enough time to complete the paper. There were the usual arithmetic and algebraic errors, but, when using the tables, the main errors were to write down the probability corresponding to an adjacent position (one place either side, up or down) of the required answer or using the wrong ‘block’, for example looking up in B(30, 0.5) instead of B(30, 0.25).

Report on Individual Questions

Question 1

Overall this question was well done.

The vast majority of students were able to gain full marks for part (a), however, any errors that arose were generally from writing $\mathrm{P}(Y \geq 10) = 1 - \mathrm{P}(Y \leq 10)$ or 1 – P(Y < 9) or simply not looking the information up correctly in the tables.

Part (b) was also very well completed. Most students were able to gain full marks here and most used the Poisson formula as opposed to the tables. The most common mistakes were finding $\mathrm{P}(Y \leq 2)$ rather than P(Y = 2), or using an incorrect value of λ. A curious error was to approximate 1.8 by 2 to allow the question to be solved by use of tables. This scored zero marks.

In part(c) most students realised the normal distribution was required and remembered that the mean should equal the variance. Problems arose in applying the continuity correction with 470 and 470.5 being popular wrong routes to take although 468 and 468.5 did make a number of appearances. Overall, a good proportion of students knew enough to use the correct continuity correction and standardisation though some found the wrong area and failed to gain the final A1.

Question 2

The response to part (a) was disappointing. A significant number of students were unable to cope with the expression $\mathrm{P}(5 \leq X < 11)$. There were students who translated this expression into the more convenient form $\mathrm{P}(5 \leq X \leq 10)$ and then in turn transformed this into an equivalent form that can be applied to the table of cumulative probabilities: $\mathrm{P}(X \leq 10) - \mathrm{P}(X \leq 4)$. However, there were also many instances of incorrect versions such as: $\mathrm{P}(X < 11) - \mathrm{P}(X \geq 5)$, $\mathrm{P}(X \leq 10) - \mathrm{P}(X \leq 5)$, $\mathrm{P}(X \leq 10) + \mathrm{P}(X \geq 5)$, $\mathrm{P}(X \leq 10) - (1 - \mathrm{P}(X \geq 5))$ and $\mathrm{P}(X \leq 11) - \text{either } \mathrm{P}(X \leq 5) \text{ or } \mathrm{P}(X \leq 4)$.

Full marks for part (b) occurred relatively rarely. Many students seemed to omit or slip up on at least one detail: forgetting the hypotheses, or using the wrong letter, or forgetting a 'conclusion in context'. Perhaps the most serious error, fortunately not too common, was the use of P(X = 1). The central feature of a hypothesis test is to establish the probability of an event 'as bad or worse than that observed' under the conditions of the null hypothesis. In this case, the relevant event is $\mathrm{P}(X \leq 1)$.

Question 3

The majority of students obtained full marks on part (a) using the standard method given in the mark scheme. A few students, some of whom were successful, used an alternative method. Anticipating part (c), they found the cumulative distribution function, F(x), in terms of k and a constant C. A minority thought that this was a discrete distribution and tried $\sum x\mathrm{P}\left(X = x\right)$. Very few students tried to 'wangle' their working out to find $k = \frac{1}{4}$.

The most successful approach to part(b) was to use a sketch; it tended to be those students who did not do a sketch who incorrectly thought the mode was 6 and with a significant number giving 1, instead of 2 (because f(2) = 1).

In part(c) most students were able to find the CDF for the first part of the function (line 2) and also correctly stated line 1 and line 4. Difficulties arose with obtaining line 3; Most of these students neither used F(6)=1 or the alternative method, just integrating the second part of the pdf hence gaining an incorrect constant.

Of those who were successful at completing this part the method that was predominantly used was the alternative method given in the markscheme   ie $\mathrm{F}\left(2\right)+\int_2^x \mathrm{f}\left(x\right)\mathrm{d}x$

It was pleasing to see that most students attempted to put their line 3 of F(x)= 0.75 for part(d), but some used the sum of lines 2 and 3 = 0.75, and others thought they still needed to integrate their line 3 of the CDF.

Many students who had an incorrect quadratic made no attempt to solve it.

Question 4

Parts (a) to (c) were routine, and most students obtained full marks. A few students integrated in (c) in contrast to the majority who used the formula $\frac{1}{12}\left(b-a\right)^2$ from the formula book

Part (d) was not technically difficult. Correct solutions were invariably brief yet fully detailed. The biggest challenge was reading and interpreting the question and deciding upon a strategy. There were in fact two equivalent strategies:

- $Y \sim \mathrm{B}\,(30{,}0.2)$then find $\mathrm{P}(X \le 3)$

- $Y \sim \mathrm{B}\,(30{,}0.8)$then find $\mathrm{P}(X \ge 4)$

The majority of students provided a correct solution to part (e). A small number of students evaluated F(0.4) or substituted 0.3 rather than 0.4; a few decided to integrate the c.d.f. and then do the substitution.

Part (f) The majority of students provided correct solutions although a significant number went down the Normal approximation route rather than using a Poisson.

Question 5

Solutions to part (a) were encouraging and showed a good knowledge and understanding of a critical region. Only a few wrote their critical region as a probability and only rarely was an interval given as between two values or the critical values on their own stated. For the students who did not give a correct critical region it is worth noting that to get the first method mark a probability from Po(4) needs to be written down as evidence of "using" Po(4). A few students were not clear about what the CR was preferring to give $0 < X < 9$ as their solution or simply stating 0 and 9.

In part (b), many students failed to give their hypotheses. If a question asks you to state your hypotheses clearly then you need to do so in the relevant part of the question to gain the mark. A number of students did not use their CR from part (a) and spent time finding another route to make the test. A conclusion in context is required, even though acceptable versions in this case could be quite minimalist such as "*Liftforall's* claim is true".

Although there were many good solutions to part(d) it required students to keep a very clear head about what they were counting. The problem was not one of subject knowledge, nor was it one of technical competence. Instead, the challenge was in interpreting the question and devising a suitable strategy. Most students were able to find the correct Poisson parameter of 6 and a correct probability. After this point the solutions were very variable. The students who wrote down either of the correct binomial distributions generally made more progress. Sometimes they calculated only one probability, normally that of paying exactly £500. There were a number who recognised that they needed two probabilities but were unable to apply the binomial formula correctly and incorrectly used 3 as their coefficient instead or 4.

Question 6

It was clear that a large majority of students found the integration very easy, but many failed to simplify powers of 1, with 1^{n+1}, 1^{n+2} and 1^{n+3} regularly seen unsimplified throughout responses. Even students who eventually simplified such expressions in part (d), usually did not go back and simplify their previous answers.

In part (a), most students realised that they had to equate $\int_0^1 kx^n \mathrm{d}x$ to 1, although the use of the limits sometimes appeared later.

Nearly all students knew that they had to find $\int_0^1 x.kx^n \mathrm{d}x$ in part (b), although a common error was claiming that $x.x^n = x^{2n}$

Students who had correctly answered parts (a) and (b) had little trouble with successfully answering part (c). However, a misunderstanding of the 1st law of indices was again apparent with some claiming that $x^2.x^n = x^{3n}$.

In part(d) a minority of students were unable to distinguish between $E(X^2)$ and Var(X). Students who immediately substituted $n = 2$ were usually able to calculate Var(X) correctly, but some relatively complex functions of x and n were seen by examiners. Most responses correctly used $\mathrm{Var}(3X) = 9\mathrm{Var}(X)$, although a surprisingly high, but significant, minority for S2 thought that $\mathrm{Var}(3X) = 3\,\mathrm{Var}(X)$.

Question 7

It was pleasing to see so many completely correct solutions to question 7. At the other extreme, there were a small number of students who made no progress beyond establishing the probabilities of each coin separately.

Then in the middle there were students who adopted a correct strategy, but failed to master all the details. The biggest single problem was the 'triple' (10, 20, 50). Some students did not realise that this has a median of 20, and omitted it from their list of triples and probabilities. Other students included this triple, but multiplied by the factor of 3, rather than 6.

Even though many students obtained full marks, some achieved this more efficiently than others. One group of students created a huge list, effectively a table with four columns: triple, median, calculation of probability of the triple, and finally the result of the calculation. There were often 27 lines in this table. Other students realised that some of the triples could be grouped together, since not only do they have the same median, they also have the same probability: for example the three triples: (10, 10, 20), (10, 20, 10) and (20, 10, 10). These students then constructed a shorter list, with 10 lines. They typically wrote, for example, (10, 10, 20) × 3. Some students went even further: when they realised that some of the calculations of the probabilities were identical, they reduced their list to just 6 lines.

The most common errors were to:

- Use the shorthand “1, 2, 5” for the medians 10, 20 and 50. Some students translated back at the end to show their results in a table with the correct medians, and no harm was done. But other students forgot to do this, and continued to work with “1, 2, 5” right to the end.

- Calculate the probability of a median of 10 only, by conventional means. They then subtracted this probability from one and divided by two to give their alleged probabilities for the medians of 20 and 50. The probabilities $P(X = 20)$ and $P(X = 50)$ are equal, but this does not then mean the probabilities for the medians of 20 and 50 are also equal.

- Students who performed calculations for all three probabilities, who then checked and discovered that these probabilities did not add up to one. A few students then decided to alter, not just the probability, but the calculation leading up to it. But they made the wrong choice: they altered a probability that was already correct, and now made it incorrect.

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