Examiners' Report

Summer 2014

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 2014
Publications Code UA040122

Mathematics Unit Statistics 2

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 well presented and although students knew the statistics needed to answer the questions they were let down by poor algebraic skills.

The main errors were to write down the probability corresponding to an adjacent position of the required answer or using the wrong 'block', for example looking up in B(30, 0.25) instead of B(25, 0.5).

Report on Individual Questions

Question 1

This was an accessible opening question. Students were confident in identifying Poisson distribution with correct parameters and were able to use it for finding the values of the probabilities required. The most common errors made by students were to work out $\mathrm{P}(X < 10)$ or $\mathrm{P}(X > 10)$ in Q01a(ii).

In Q01(b) λ was usually calculated correctly for the different time period. However many students did not understand that to work out the probability that "the next patient arrives" is the same as $1 - \mathrm{P}(X = 0)$. Many calculated $\mathrm{P}(X = 1)$ orP(X£1).

Question 2

This question differentiated very well between students.

Most students produced a fully correct solution in Q02(a). However, a significant minority failed to show where the factor of 486 came from, and were penalised 2 marks. Nearly all students equated their definite integral to 1 at a sufficiently early stage to convince examiners that the result $c = \frac{1}{486}$ had genuinely been obtained. Very few cases of verification were seen, though these were usually done sufficiently well to secure full marks. A few tried to treat this as a discrete distribution substituting every integer from 1 to 9 into the formula and adding them to = 1.

Q02(b) produced a mixed response. There were many perfect solutions, which mainly tended to use definite integration rather than evaluating the constant of integration. However, a significant minority of students neither used limits nor a constant of integration, and consequently gained no credit. Some students failed to completely define $\mathrm{F}(t)$, as given in the question, and so lost the second mark.

In Q02(c) a significant majority scored full marks by evaluating $1 - \mathrm{F}(3)$. Other students worked from first principles and integrated the pdf between the limits of 3 and 9. The most common errors were to use the limits 4 and 9 instead of 3 and 9 and to find either F(3) or 1 - F(2).

Q02(d) proved to be very challenging for the majority of students. The conditional probability eluded many students, and so they were unable to secure any marks. It was not uncommon to see $\mathrm{P}\left(T > 7\right) = 0.068587$ ... given as the final answer, whilst others multiplied á $\mathrm{P}(T > 3)$ by $\mathrm{P}(T > 7)$. When a quotient of probabilities was given, the two most common errors were giving an incorrect numerator and using F(3) as the denominator. However, a small minority of students were able to complete this part correctly.

In Q02(e) the most common errors were simply to square the answer to Q02(c) and the omission of the factor of 3.

Question 3

In Q03(a) many students had learnt when a Poisson distribution is suitable but many did not use the context given in the question. It was pleasing to see that few students gave independent and random as separate reasons although some students lost the mark for "constant rate" by talking about rate or giving the conditions needed when approximating to a normal distribution.

Most students knew how to set about finding a critical region in Q02(b) and located the correct probabilities but gave the incorrect critical regions with $X \geq 8$ being the most common error although $X \geq 10$ and $X \geq 7$ were occasionally seen. A few students gave the critical region using probability statements although a small minority gave critical values rather critical regions.

Q03(c) was usually correct even if their answers to Q03(b) had failed to gain full credit.

In Q03(d) the majority of students stated whether 8 was or was not in their critical region and drew the correct contextual conclusion and managed to associate the "claim" with H_0 rather than H_1. A few had not appreciated the connection with Q03(b) and redid the entire calculation. A minority of students did not appreciate that a statistical justification was required and said that 8 was double the expected number of emails and so was a significant result.

Q03(e) was well done, even by students who had gained few marks up to this point. The first 3 marks were usually scored, although calculating $P(X = 2)$ or $P(X \leq 1)$ was sometimes seen. A minority wrongly accepted H_0 and did not include enough context in their conclusion. An error sometimes seen was to compare the probability with 0.05 and incorrectly accept the null hypothesis.

Question 4

The majority of students were able to use the correct model in Q04(a)(i) and calculated $\mathrm{P}(X = 6)$ correctly. In Q04(a)(ii) many students found the correct answer but not always by the quickest method which was using $\mathrm{B}(10, 0.25)$ and finding $\mathrm{P}(X \le 2)$. Others found $\mathrm{P}(8) + \mathrm{P}(9) + \mathrm{P}(10)$ and used the formula for $\mathrm{B}(10, 0.75)$ or even used the formula to calculate $\mathrm{P}(X \le 7)$ and then subtract from 1. Some students incorrectly changed their model to $\mathrm{Po}(7.5)$.

Q04(b) was frequently solved efficiently and completely correctly. However, a significant minority of students did not recognise that they should take $0.2^{1/20}$ to obtain $1 - p$ instead choosing to use the tables to obtain an estimate for p, usually obtaining 0.1. Other students incorrectly thought that 0.8 (or 0.2) was the value of p and not $1 - \mathrm{P}(X = 0)$.

Q04(c) proved to be challenging. Most students were able to see that the exact distribution was B(100, 0.975) for the number of times the cadet hits the target. Few realised that they should define B(100,0.025) as the model for the number of times the cadet misses the target and then follow this with an approximation to the Poisson obtaining $\mathrm{Po}(2.5)$. Even when they used $\mathrm{Po}(2.5)$ many were unable to change " the cadet hits at least 95 " to " the cadet misses at most 5 ". Those that stayed with B(100, 0.975) approximated to the normal, saying things like "mean = 97.5 this is too big for the tables so approximate to the normal".

Question 5

Correct responses to Q05(a) were seen from the majority of students, although p "small" was sometimes offered. Students who stated that $np > 5, nq > 5$ usually did so in addition to, rather than instead of, the only accepted response of n is large and p is close to 0.5.

Q05(b) elicited a mixed response, ranging from a null response, through "too time consuming and too expensive" to the accepted answer that there would be no pea seeds left to sell. The words "it" and "they" sometimes were used without any reference as to context.

The correct hypotheses were correctly given by a large majority of students in Q05(c). Sometimes they appeared in Q05(d), which was accepted, but occasionally different hypotheses appeared in Q05(c) and Q05(d). Occasionally either a letter other than p was used or hypotheses for a 1-tailed test were given.

In Q05(d) a large majority of students correctly calculated the values of the two parameters of the normal distribution, although 121 sometimes appeared as both the mean and variance following a normal approximation to the Poisson distribution Po(121).

The application of a continuity correction caused problems to many students; some were unaware that one should be used, whilst others used 135.5 instead of 134.5. Students ought to be made aware that the omission of a continuity correction resulted in the loss of at least 4 marks. A significant minority of students were able to obtain 0.0336. Few students chose to find the upper critical value. The standardisation was virtually always correctly done with the students' values.

Although nearly all students were using a 2-tailed test, it was not uncommon to see students comparing 0.0336 with 0.05 (instead of 0.025), and hence arriving at the wrong conclusion. Nevertheless, a large majority of students knew that they were expected to give a statement and contextual conclusion. The latter was sometimes marred by a failure to give a complete conclusion; in particular, the words "company's" or "seeds" were sometimes omitted.

Question 6

Students used accurate and detailed working leading to a correct answer in Q06(a). One relatively common incorrect method involved integrating the pdf (as opposed to $x\mathrm{f}(x)$). Many of these students ended up with a final answer of 1.

The most common errors were to use only one of the three 'parts' of the pdf, add all three 'parts' before integrating, and integrating all three 'parts' successfully, but then dividing by three.

Q06(b) was not technically difficult: but persistence and attention to detail were required, particularly regarding algebra and arithmetic. However, the real problem for many students was conceptual. Almost all students could integrate all three parts of $\mathrm{f}(x)$. However, a substantial number apparently did not address this central notion of cumulative probability: they failed to add $\frac{1}{9}$ for the second part and $\frac{7}{9}$ for the third part. The students who used the $+C$ method were able to use $\mathrm{F}(0) = 0$ and $\mathrm{F}(6) = 1$ to gain the 2nd and 4th line of the c.d.f however they did not know what to use to calculate the C to gain the 3rd line; many simply used $\mathrm{F}(0) = 0$ for a second time.

It was noted that almost all students adopted the correct general strategy for Q06(c), which is to solve the equation $\mathrm{F}(m) = 0.5$. On this occasion, only a minority of students chose the incorrect method of evaluating $\mathrm{F}(0.5)$. However, a significant number of students used the wrong part of their version of $\mathrm{F}(x)$. Students would be advised to do a little bit of preliminary work, which consists of: $\mathrm{F}(1) = \frac{1}{9} < \frac{1}{2}$; $\mathrm{F}(4) = \frac{7}{9} > \frac{1}{2}$ so $1 < m < 4$ so use the third line of $\mathrm{F}(x)$

In Q06(d) very few students said there was little skewness or symmetrical.

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