Examiners' Report/
Principal Examiner Feedback

Summer 2012

GCE Statistics S2 (6684) Paper 01

Edexcel and BTEC Qualifications

Edexcel and BTEC qualifications come from Pearson, the world's leading learning company. 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 for our BTEC qualifications.
Alternatively, you can get in touch with us using the details on our contact us page at www.edexcel.com/contactus.

If you have any subject specific questions about this specification that require the help of a subject specialist, you can speak directly to the subject team at Pearson.
Their contact details can be found on this link: www.edexcel.com/teachingservices.

You can also use our online Ask the Expert service at www.edexcel.com/ask. You will need an Edexcel username and password to access this service.

Pearson: helping people progress, everywhere

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 2012
Publications Code UA033139

Introduction

The paper proved accessible to the majority of candidates and there was little evidence of there not being enough time to complete the paper. There were fewer arithmetical errors than in previous examinations, but candidates need to take greater care when using statistical 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, 0.25) instead of B(25, 0.5). The presentation of the work was generally good with candidates showing their methods clearly.

Report on individual questions

Question 1

This question was accessible to the majority of candidates, with many gaining full marks. Most were familiar with the continuous uniform distribution and were able to find $P(L > 24)$ in part (a) A small minority of candidates used a discrete uniform distribution and calculated $P(L \geq 25)$ giving a probability of 1/3. Part (b) was generally well answered and candidates used binomial tables for B(20, 0.4) with few errors. The most common error was to write and then evaluate $P(X \geq 8) = 1 - P(X \leq 8)$. Candidates who got an answer of 1/3 for part (a) struggled with using B(20, 1/3) in part (b) and resorted to finding $P(X = 8)$. A small minority used Poisson or normal distributions e.g. Po(8), N(20, 0.4) or N(40, 0.4). Those who gained few marks in parts (a) and (b) were often able to gain obtain full marks for part (c) for finding the square of their answer to (b). The most common errors in part (c) were to double the answer in (b), rather than square it or use the distribution $\mathrm{B}(40,0.4)$ and evaluate $\mathrm{P}(X \leq 8)$.

Question 2

Part (a) was a routine question with many fully correct answers being seen. However, a minority of candidates were unable to write down the critical regions correctly. The two most common errors were to write down the regions in terms of probabilities e.g $\mathrm{P}(X \leq 7)$ or to simply write $X = 7$ and $X = 18$.

In part (b) many candidates were distracted by the word "incorrectly". We reject the null hypothesis 'incorrectly' when it is in fact true. In this case, this means that p really is 0.5, so that we can use the distribution B(25, 0.5) to work out probabilities. This is the distribution that had already been used in part (a), so that all that needed to be done was to simply add the two probabilities that were used to identify the two parts of the critical region in (a): 0.0216 + 0.0216 = 0.0432. It was all too common to see correct working followed by incorrect and irrelevant work that invalidated the whole response. The most common incorrect postscripts were:

- 0.05 − 0.0432 = 0.0068
- 1 − 0.0432 = 0.9568

Question 3

This question was accessible to the majority of candidates. Whilst many candidates knew "the two conditions needed to approximate the binomial distribution by the Poisson distribution this knowledge was by no means universal. Some candidates appear to have tried to memorise the conditions for:

- modelling a situation using a Binomial distribution
- modelling a situation using a Poisson distribution
- approximating the binomial distribution with the Poisson distribution
- approximating the binomial distribution with the Normal distribution

but failed to remember which set of conditions applies in which situation.

The response to part (b) was very good, with many candidates gaining full marks. The most common error made was to state the hypotheses as $\mathrm{H}_0 : \lambda = 6$ and $\mathrm{H}_1 : \lambda > 6$. The question clearly states that we are testing for a 'proportion', so that the null hypotheses should be $\mathrm{H}_0 : p = 0.03$ and $\mathrm{H}_1 : p > 0.03$. The majority of candidates used the correct Poisson distribution and successfully calculated the probability 0.0201. It was common to see candidates either using the incorrect statement P(X ≥ 12) = 1 – P(X ≤ 12) or calculating 1 – P(X ≤ 10) and writing CR: X ≥ 10. The candidates who tried to find the critical region were more likely to make an error and it is recommended that the probability route is used.

It is pleasing to see that most candidates finished their answers well with a clear conclusion using the context written in the question.

Question 4

A high proportion of the candidates attempted all parts of this question successfully. Those less successful in part (a) either used Po(2) instead of Po(8), or used Po(8) but found the answer, using Poisson tables, to P(X ≤ 3) instead of P(X = 3). A common error seen in part (a)(ii) was to write and use P(X > 5) = 1 – P(X ≤ 4).

In part (b) many candidates were able to find P(Y = 9) using their answer to (a)(ii) but, for a small minority, incorrect calculations included calculating $(0.8088)^9$ or using Po(λ) with λ = 8, 9 or 24. Overall, many exemplary answers were evident for the solution to part (c) reflecting sound preparation on this topic. Marks were lost occasionally for using an incorrect, or no, continuity correction or for finding an incorrect area.

Question 5

Many correct responses to part (a) were seen with a large majority of candidates choosing to use the method of solving the equation $\int_0^k \mathrm{f}(x)\,\mathrm{d}x = 1$ for *k*.

A small number of candidates successfully 'verified' that $k = 4$ by substituting $k = 4$ and then showing that $\int_0^4 \mathrm{f}(x)\,\mathrm{d}x$ is in fact 1.

The main errors were
(i) a partial verification attempt in which it was assumed that *k* = 4 in the integrand only, and then integrating to "verify" that *k* = 4;
(ii) extracting a "factor" of *k* from the integral;
(iii) inability to multiply out the brackets and then forming the product of the integrals of the two factors.

In part (b) few candidates appeared to notice the use of symmetry. The majority of candidates evaluated E(*X*) by integration, with some noting that "E(*X*) =2 by symmetry" at the end of their calculation. The instruction "Write down", as well as the allocation of just 1 mark should have alerted candidates that a calculation was unnecessary here.

Part (c) was extremely well answered. There was much evidence of confident, fluent and correct techniques of algebra and calculus with most candidates scoring full marks, although a small proportion did not subtract $[\mathrm{E}(X)]^2$ from $\mathrm{E}(X^2)$. There were also a small number of candidates who attempted to use the correct alternative method mentioned in the formula booklet: $\int (x-\mu)^2 \mathrm{f}(x)\,\mathrm{d}x$. This approach is generally not to be favoured as it is much less convenient than the standard method: $\int x^2 \mathrm{f}(x)\,\mathrm{d}x - \mu^2$. The candidates who attempted this alternative method were generally not successful and it is suggested that this method is best avoided.

Part (d) proved to be a test of candidates' linguistic skills, with many candidates calculating the complementary probability. Stronger candidates produced relatively concise responses using symmetry. The use of incorrect limits such as 0 and 0.5 or 0.5 and 2 was seen several times, and a few candidates confused f(*x*) with F(*x*).

Question 6

Virtually all candidates were able to write down the eight combinations of 1s and 2s and although most candidates were able to calculate the probability of each of these outcomes, the meaning of "range" seems to have eluded all but a minority of candidates. The key to the question was remembering that the 'range' of a list of numbers is obtained by subtracting the lowest number from the highest.

Many candidates obtained the sampling distribution of (in descending order of popularity) the mean, the total score, the median or the mode. Some candidates gave the sampling distribution of two (or more) of these statistics, without ever considering the range.

Of those candidates who considered the range, the most common error was to state that the range was either 1 or 2.

Question 7

Although a reasonable proportion of candidates produced a correct sketch in part (a), many were unable to do so. Each of the three sections of the sketch required no more than GCSE Mathematics. Despite this,
(i) $f(x) = x^2/45$ was frequently identified as a straight line or sometimes as a curve, but with incorrect curvature;
(ii) $f(x) = 1/5$ sometimes appeared as a sloping straight line or a curve;
(iii) $f(x) = 1/3 - x/30$ was often interpreted as a curve.

There were many excellent responses to part (b), but misunderstanding of the cdf was widespread. Most candidates managed to obtain the equation of the first section as $x^3/135,\ 0 \le x \le 3$, but evaluation of the constant term in each of the next two sections often proved difficult. The two approaches that were used are

- indefinite integration using "+ c"
- definite integration and then adding on the area of any earlier sections

The most common error for those who selected the first method was to use an indefinite integral without a constant of integration. Those who had a constant of integration were often unsuccessful because they believed that $3^3 = 9$.
Interestingly, some candidates who had failed to obtain the correct equation for the 2nd section, then went on to successfully find the equation for the 3rd section using $F(10) = 1$ to evaluate their constant of integration.
For the candidates who used definite integrals the main error was to forgot to add on $F(3)$ for the 2nd section and $F(4)$ for the 3rd section.

Many candidates correctly wrote down $F(x) = 0,\ x < 0$ and $F(x) = 1,\ x > 10$, though carelessness was occasionally evident, e.g. $F(x) = 0,\ 0 < x$. Other candidates seemed to be confusing the pdf with the cdf by writing $F(x) = 0$, otherwise.

The evaluation of $F(8)$ in part (c) was often attempted successfully by candidates, some only obtaining the method mark because of an incorrect expression for $F(x)$. Various other valid methods, based upon areas under the pdf, were also accepted.

It is worth noting that some candidates evaluated f(8) rather than F(8), gaining no credit. Others treated X as a discrete random variable, evaluating 1 – [P(X = 9) + P(X = 10)].

Question 8

There were many correct answers to part (a). However, this was not a unanimous response. The theme of this question was probability distributions that do not appear in the tables. There were many candidates whose strategy was to work around this problem by using a Poisson approximation to the Binomial. For this particular Binomial distribution, $X \sim \mathrm{B}(10, 0.6)$, it is not appropriate to use a Poisson approximation. The candidates who were confident using the formula for Binomial probability were most successful, obtaining the required probabilities in (i) and (ii) without the use of the cumulative probability tables. The candidates who felt they must use the tables and considered the related distribution $Y \sim \mathrm{B}(10,0.4)$ were not quite as successful. Not all candidates made the necessary amendments: for instance, in (a)(ii) we require $\mathrm{P}(X \le 8)$, but cannot just look up $\mathrm{P}(Y \le 8)$ in the tables. It is essential to realise that X and Y are related by $X + Y = 10$. We can therefore replace X by 10 – Y, and rewrite $\mathrm{P}(X \le 8)$ as $\mathrm{P}(Y \ge 2)$, and then evaluate this using the standard strategy of $1 - \mathrm{P}(Y \le 1)$.

In part (b) many candidates appreciated that they needed to use either B(50, 0.6) or B(50, 0.4). However, many were unable to proceed further. Success in part (b) relied on using the relationship $X + Y = 50$ so that X can be replaced by 50 – Y. After some algebraic manipulation, the inequality in the question, $\mathrm{P}(X < n) \ge 0.9$, is transformed into $\mathrm{P}(Y \le 50 - n) \le 0.1$. This proved elusive to the great majority of candidates and the final two marks were often not awarded. Very common incorrect answers were 15, 16, 24, 25, 26 and 34.

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

Further copies of this publication are available from
Edexcel Publications, Adamsway, Mansfield, Notts, NG18 4FN

Telephone 01623 467467
Fax 01623 450481
Email publication.orders@edexcel.com

Order Code UA033139 Summer 2012

For more information on Edexcel qualifications, please visit
www.edexcel.com/quals

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

Ofqual

Llywodraeth Cynulliad Cymru
Welsh Assembly Government

cea
Rewarding Learning