edexcel

Examiners' Report/
Principal Examiner Feedback

June 2011

GCE Statistics S2 (6684) Paper 1

Edexcel is one of the leading examining and awarding bodies in the UK and throughout the world. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers.

Through a network of UK and overseas offices, Edexcel's centres receive the support they need to help them deliver their education and training programmes to learners.

For further information, please call our GCE line on 0844 576 0025, or visit our website at www.edexcel.com.

If you have any subject specific questions about the content of this Examiners' Report that require the help of a subject specialist, you may find our **Ask The Expert** email service helpful.

Ask The Expert can be accessed online at the following link:
http://www.edexcel.com/Aboutus/contact-us/

June 2011
Publications Code UA028839

Statistics Unit S2
Specification 6684

Introduction

Candidates would appear to have run short on time for this paper. This was perhaps due to them not using the most efficient methods to answer questions such as question 3. There were few times when no attempt had been made to produce an answer to a question.

The level of work was very mixed. The standard of presentation was weak – it was often very difficult to follow the work and different parts were not clearly labelled. Presentation of working on a page was often disorganised, this was particularly the case on questions 3 and 7.

Report on individual questions

Question 1

In part (a) many candidates had learnt standard definitions but were generally unable to apply these definitions in context. Although the word 'list' was evident in the majority of answers it often referred to the components or the sample and not the identity numbers.

Many candidates failed to appreciate what was required in part (b), merely stating the standard requirements of a sampling distribution i.e. all the values taken by the variate and their associated probabilities. Indeed in some cases it was appreciated that B(50,0.02) was required, but unfortunately this appeared in part (a) and went unrewarded. A small number of candidates took 2% as 0.2.

Question 2

There was a very variable response to question 2, with many candidates producing "textbook answers", whilst many others failing to recognise a Poisson distribution in part (a), offered either (or sometimes both) a binomial or a normal model.

The latter candidates either stopped at part (a) or pursued their chosen model to little effect.

In part (b) the vast majority successfully opted for a 1-tailed alternative hypothesis, although some did insist on using the parameter p. The value of $P(X \geq 12)$ or the CR was usually found correctly and most candidates were able to make a successful comparison, thereby leading to a well expressed contextual conclusion. Some candidates whose alternative hypothesis suggested a 2-tailed test, still opted to perform a 1-tailed test.

Question 3

Many candidates found this question challenging and there were few fully correct answers. Part (a) was well answered.

In part (b), almost all candidates correctly integrated the given pdf, between the limits of 0 and 3. Unfortunately, the majority equated their definite integral to 1, instead of $\frac{1}{2}$ showing a misunderstanding of the concept that the total area is equal to 1.

In part (c), equating the area of the triangle to $\frac{1}{2}$ proved to be surprisingly demanding for many candidates, although there were a significant percentage of concise solutions. A minority opted for finding the equation of the line segment, before integrating it. This in itself was rather complex especially if they chose B as the point on the line rather than (*a*, 0). This then had to be integrated leading to a lot of complex algebra, which rarely resulted in a correct solution. Some candidates clearly guessed the answer to (a) by looking at the diagram - this was awarded no marks.

Part (d) was very poorly answered. Only a small percentage of candidates stated that the distribution was negatively skewed. The vast majority were under the illusion that as median = mode then the mean must be the same value thus drawing the conclusion that it was symmetrical, despite evidence to the contrary on the diagram drawn on the question paper.

Question 4

Part (a) was routine and the vast majority of candidates demonstrated familiarity with the probability density function of a rectangular distribution.

Part (b) required some careful initial thought that eluded a large majority of the candidates. The successful solutions fell into two camps. On the one hand were those candidates who identified the Binomial distribution $X \sim \mathrm{B}\left(3, \frac{5}{6}\right)$ and then evaluated the probability $\mathrm{P}(X \geq 1)$ by conventional means. On the other hand there were candidates who worked 'from first principles' and produced either an elaborate tree diagram or a list of all possible outcomes. This latter approach ('from first principles') is of course valid, but takes longer and is more susceptible to error.

There were many complete and correct solutions to part (c). Finding the length of a stick shorter than 7.6 cm was straightforward for most but many attempts reflected candidates' lack of understanding of what the question required of them in this part. Successful candidates realised it was a binomial situation using B(6, p) and used tables to find $1-\mathrm{P}(X \leq 4)$ or calculated P(X = 5) + P(X = 6). Common errors included misinterpretation of P(X > 4) as $1-\mathrm{P}(X \leq 3)$ or even P(X = 4) and a small minority of candidates gave the answer as $\left(\frac{1}{5}\right)^4 = \frac{1}{625}$ which gained no marks as it is an incorrect method.

Question 5

This question was well answered with many candidates gaining full marks. Part (b) required considerable attention to detail, and it is commendable that so many candidates were able to achieve all 5 marks. There were a small number of candidates who earned four of these five marks despite an incorrect answer to part (a). These candidates demonstrated the importance of showing one's method and including all relevant details. It is often not possible to earn 'follow through' marks unless the method and necessary detail is made explicit. The most common errors were to muddle up p and q or forget the 6 in the calculation of P(X = 1).

Candidates generally gave clear, confident and accurate responses to part (c) that demonstrated mastery of both theory and detail. Although there were the usual errors of not using a continuity correction, using the wrong continuity correction or using the wrong mean and variance there were many fully correct solutions.

Question 6

This appeared to be generally a fairly straightforward question and there were many fully correct and clear solutions. In part (a) a number of candidates chose the wrong $\mathrm{H}_1 : p < 0.15$. There are two main methods for conducting this significance test. It was noticeable that the candidates who chose the 'Critical Region' approach were generally less successful than those who evaluated the probability of obtaining an outcome 'as bad or worse' than that observed. It is advised that using the Critical Region method for hypothesis tests should be discouraged. The main errors generally involved comparing the probability found with an incorrect value. It was not uncommon to see the correct probability of 0.0480 compared with 0.05, instead of with 0.025 when a two tail test had been indicated by the hypotheses. Others incorrectly found P(X = 1).

Part (b) appeared to be a lot easier than part (a) and many were able to accurately find the p value and gave good conclusions in context, Some candidates approximated to the Poisson and then to the normal so having the wrong variance for their distribution. Whilst most candidates attempted the continuity correction a substantial number wrongly used 31.5 instead of 30.5. A very small minority of the candidates attempted to use a 'Critical Region' approach however these were mostly incorrect and gained few marks and were not as successful as those who calculated a probability.

Question 7

The overall response to part (a) was disappointing. The shape of the graph is assumed knowledge from GCSE Mathematics. If $f(x)$ is a quadratic expression, then the graph of $y = f(x)$ is a parabola, whose shape is 'known'. Many candidates made a 'table of values' which they then used to plot the graph which then consisted of straight lines. Occasionally the correct parabola shape was seen with 1 and 5 marked on the *x*-axis but the graph continued below the *x*-axis indicating that the probability could be negative for values outside of the given range.

Although part (b) was intended to be straightforward, many candidates failed to see the symmetry of the pdf (even with a correct diagram), and pursued integration to find the mean. Fortunately, most attempts were successful. It should be noted that in general if there is 1 mark for a question the answer should be able to be written down without any long calculations.

Part (c) was well answered by the majority of candidates, with responses ranging from the very concise to lengthy. Nearly all candidates were able to multiply the brackets and integrate successfully, although there were occasional sign errors. Sufficient working for the substitution of the limits was nearly always shown. Some candidates preferred to find Var(X) first, only to add [E(X)]^2 at the end. Some candidates obtained an answer of almost 9.8. This is despite the fact that most modern calculators will work using exact fractions. It was in fact possible to work with exact decimals, the penultimate stage of working being $9.765625 - (-0.034375)$. However, some candidates chose to approximate the earlier decimals, resulting in a final answer that differed from 9.8.

Part (d) was usually answered correctly, with a significant minority giving their answer in surd form.

In part (e) the obvious methods, in this case solving either F(1) = 0 or F(5) = 1, were not universally adopted. For those who adopted the latter approach, a surprising proportion struggled to solve the linear equation: $\frac{1}{32}(a - 75 + 225 - 125) = 1$. Many candidates integrated f(x), or occasionally F(x), with mixed success.

The solution to part (f) had been well rehearsed by candidates, although some failed to give an adequate explanation for the final mark. An ideal correct response such as "0.25 is between 0.245 and 0.252, so the median must lie between 2.29 and 2.31" was rarely seen. A sizeable minority pursued the solution of F(x) = $\frac{1}{4}$. Those who used a sign change method were sometimes successful, and a few candidates used their graphical calculators to state all three roots, before selecting the required one.

In part (g), many candidates did not appreciate the symmetric nature of the pdf. However, those who did usually produced a correct answer of 3.7. Other candidates resorted to a repetition of their method for finding the lower quartile.

In part (h) those candidates who saw that the given probability referred to the IQR of the distribution often produced a neat solution. More often, however, candidates abandoned their solution, usually after much algebraic effort.

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 UA028839 June 2011

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

Rewarding Learning