Examiners' Report

January 2010

GCE

Statistics S2 (6684)

Edexcel Limited. Registered in England and Wales No. 4496750
Registered Office: One90 High Holborn, London WC1V 7BH

A PEARSON COMPANY

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, our GCSE team on 0844 576 0027, 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/

January 2010

Publications Code UA023028

Statistics Unit S2
Specification 6684

Introduction

Candidates would appear to have had enough time to complete this paper. There were few questions where no attempt had been made to produce an answer.

The level of work was generally good. Many candidates were well prepared for the examination and candidate's answers to Q4 and Q7 were better presented and answered than in previous years

Report on individual questions

Question 1

This question proved to be a very good start to the paper for a large majority of candidates. In general parts (a), (b) and (d) were answered correctly. In part (c) the most common mistake was to use $P(X > 4) = 1 - P(X \leq 3)$.

Question 2

Most of candidates were able to gain the majority of the marks for this question. In part (a) most candidates either correctly substituted 0 into the formula for $F(x)$ or used $F(0) - F(-2)$.
A common error was to integrate $F(X)$, which in many cases resulted in a correct answer but gained no marks as the method was incorrect. There were also a number of students who believed the distribution to be discrete and calculated $F(x)$ accordingly.

In part (b) there were a few candidates who integrated $F(x)$ or used $F(x)$ rather than differentiating to find $f(x)$. Some of those who differentiated correctly then failed to identify the regions in which the values of $\frac{1}{6}$ and 0 were valid.

In part (c) a large number of candidates were unable to completely state the name of the distribution with common errors being to omit either the word 'continuous' or the word 'uniform'. In part (d) although most candidates were able to identify or calculate the mean, a few carried out complicated unnecessary calculations, which were usually incorrect. The candidates that used the formula usually achieved a correct solution for the variance. However, those that attempted to use $\int x^2 f(x)$ - mean^2 often made errors or forgot to subtract the mean squared.

In part (d) many candidates failed to realise that the probability that X equals a single value in a continuous distribution is always 0.

Question 3

Although there were a minority of candidates who were unable to identify the correct distribution to use the majority of candidates achieved full marks to parts (a) (b) and (c). Part (d) seemed to cause substantial difficulty. In part (a) the majority of candidates identified that a Poisson (rather than the Binomial) distribution was appropriate but some calculated the parameter as 2.5 or 4 rather than 0.4. A few used Po(1) and calculated P(Y=5).

In part (b) and part (c) the most common error was to use Po(2.5). The majority of candidates were able to work out P(X >1) and P(X = 2) using the correct Poisson formula. Many thought that their answer to part (c) was the correct solution while others used or multiplied their answers to both (b) and (c). Whether stating a correct or incorrect solution only a minority used the statistical term "independence" as the reason for their answer.

Question 4

Many candidates scored full marks in part (a). However, there was inevitably substantial variation in the style and presentation of their arguments. A small number of scripts were models of clarity and economy, while other candidates were not just long-winded and confusing, but their scripts contained incorrect and contradictory statements (e.g. $6k = 1$, $3k = 1$, $9k = 1$) on the way to a correct final solution.
It is reassuring that many perfect solutions to part (b) were seen. The most common problem was the part of the distribution dealing with $3 \le x \le 4$ where many candidates simply worked out $\int_3^x \frac{1}{3}\,\mathrm{d}t = \frac{x}{3} - 1$. Of those who got the correct answer a mixture of methods were used.

Some use the approach $\int_3^x \frac{1}{3}\,\mathrm{d}t + \mathrm{F}(3)$ where $\mathrm{F}(3) = \frac{2}{3}$, while others chose indefinite integration: $\int \frac{1}{3}\,\mathrm{d}x = \frac{x}{3} + \mathrm{C}$ and $\mathrm{F}(4) = 1$.

In part (c) correct answers seemed relatively elusive. A significant number of candidates attempted to find E(X) by using $\int x\mathrm{f}(x)\mathrm{d}x$ for one part only, usually the f(x) for $3 \le x \le 4$. There were a small number of candidates who 'averaged' their answers to the two parts: $\frac{5/4 + 7/6}{2}$. Other candidates multiplied the F(x) by x before integrating.

Part (d) was well done by many candidates, even when there had been problems earlier in the question. Most understood what they were trying to do, and often had a correct version of the function to hand. However, some failed to provide an acceptable conclusion: "so the median is between the two numbers" or "F(2.6) < m < F(2.7)" are examples which did not gain the mark.
The most common error was to amalgamate their two functions from part (b). Others used the 'wrong' function, i.e. their function from (b) for $3 \le x \le 4$.
Those candidates who used calculators to solve the cubic equation usually provided the required amount of supporting detail.

Question 5

This question was accessible to the majority of candidates, with many gaining full marks. Most recognised the need to use a Poisson distribution in part (a) and translated the time of one hour successfully to a mean of 10. Common errors included using a mean of 6 or misinterpreting $P(X < 9)$ as $P(X \leq 9)$ or using $1 - P(X \leq 8)$. In part (b), a high percentage of candidates gained full marks for using a Normal approximation with correct working. Marks lost in this part were mainly due to using a 49.5 instead of 50.5 or no continuity correction at all. A small number of candidates wrote the distribution as B(240, 1/6) and translated this to N(40, 100/3).

Question 6

Part (a) tested candidates' understanding of the critical region of a test statistic and responses were very varied, with many giving answers in terms of a 'region' or 'area' and making no reference to the null hypothesis or the test being significant. Many candidates lost at least one mark in part (b), either through not showing the working to get the probability for the upper critical value, i.e. $1 - P(X \leq 15) = P(X \geq 16) = 0.0064$, or by not showing any results that indicated that they had used B(30, 0.3) and just writing down the critical regions, often incorrectly. A minority of candidates still write their critical regions in terms of probabilities and lose the final two marks. Responses in part (c) were generally good with the majority of candidates making a comment about the observed value and their critical region. A small percentage of responses contained contradictory statements.

Question 7

A high proportion of candidates attempted the first two parts of this question successfully, with the majority of candidates getting at least one mark for part (b). Those less successful in part (a) either misread the question and ended up with a denominator of 3 for the probabilities or confused formulae for calculating the mean and variance and used, for example, $\sum \frac{xp(x)}{n}$ for the mean or used $E(X^2)$ for σ^2. The solution to part (c) proved beyond the capability of a minority of candidates but, for the majority, many exemplary answers were evident, reflecting sound preparation on this topic. Candidates who found all 8 cases in (b) usually gained four marks in part (c) for calculating the probabilities. For a small percentage of those candidates, calculating the means was difficult and hence completing the table correctly was not possible. A few candidates tried unsuccessfully to use the binomial to answer part (c).

Grade Boundaries

The table below gives the lowest raw marks for the award of the stated uniform marks (UMS).

Module	80	70	60	50	40
6663 Core Mathematics C1	63	54	46	38	30
6664 Core Mathematics C2	54	47	40	33	27
6665 Core Mathematics C3	59	52	45	39	33
6666 Core Mathematics C4	61	53	46	39	32
6667 Further Pure Mathematics FP1	64	56	49	42	35
6674 Further Pure Mathematics FP1 (legacy)	62	54	46	39	32
6675 Further Pure Mathematics FP2 (legacy)	52	46	40	35	30
6676 Further Pure Mathematics FP3 (legacy)	59	52	45	38	32
6677 Mechanics M1	61	53	45	38	31
6678 Mechanics M2	53	46	39	33	27
6679 Mechanics M3	57	51	45	40	35
6683 Statistics S1	65	58	51	45	39
6684 Statistics S2	65	57	50	43	36
6689 Decision Maths D1	67	61	55	49	44

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

Telephone 01623 467467
Fax 01623 450481

Email publications@linneydirect.com

Order Code UA023028 January 2010

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

Edexcel Limited. Registered in England and Wales no.4496750
Registered Office: One90 High Holborn, London, WC1V 7BH