Pearson

Examiners' Report
Principal Examiner Feedback

January 2018

Pearson Edexcel International A Level
In Statistics 2 (WST02)

edexcel

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

January 2018
Publications Code WST02_01_ER

IAL Mathematics Unit Statistics S2

Specification WST02/01

General introduction

The paper was accessible to all candidates and overall there were some strong performances. Candidates sometimes struggle when translating certain topics in context into correct probability statements, particularly the continuous uniform distribution. Conditional probability questions continue to challenge candidates with Qu 1(b) and Qu. 5(d) discriminating the most able. The quality of the sketches of the probability density function in Qu 7(b) was generally poor. It was pleasing to see that the conclusions to the hypothesis test in Qu (4) were generally given in context.

Report on Individual Questions

Question 1

The first question on the paper proved accessible to all candidates and over 40% went on to achieve full marks here. Virtually all candidates scored both marks in part (a) by correctly substituting directly into the cumulative distribution function. There were, however, some occasions where integration of $F(x)$ was seen.

Correct solutions to part (b) were seen, but these were not less common. Conditional probability is often a serious problem for a substantial number of candidates. The problem for many candidates in a standard question like this is to obtain a simplified expression for the numerator: $P((X > 3) \cap (2 < X < 4))$. The most common error here was to simplify this to $P(X > 3)$ leading to the incorrect conditional probability $\frac{1 - F(3)}{F(4) - F(2)}$. It should have been clear to candidates that this was incorrect as it gave a probability which was greater than 1.

Part (c) was well answered with many fully correct solutions seen. Some of the errors seen were fairly predictable, but fortunately not too common. Some candidates started by multiplying the function given in the question by x and then integrating. Other candidates did obtain the correct probability density function but then failed to multiply by x before integrating.

Question 2

Question 2 was a significantly more challenging question with 25% of candidates scoring 0 marks here, however, over 1/3 of candidates achieved full marks. For most candidates part (a) was a straightforward part to the question. Candidates who realised that $P(X = 6) = \frac{n}{n+1}$ generally managed to show that $n = 9$. Some candidates attempted $\sqrt[3]{0.729} = 0.9$ but did not make the link with n and simply wrote down the given answer. There were a significant minority of candidates who wrote nothing in this part.

Responses to part (b) were helped by the fact that $P(T = 18) = 0.729$ had been given and many candidates were able to find the correct sample space for T with accurate associated probabilities. A small proportion of candidates omitted "$\times$ 3" for the probabilities for $T = 24$ and 30. This should be an easy error to spot: the probabilities do not then add up to 1 - an obvious method of double-checking an answer.

Part (c) proved challenging for some candidates. The concept of the range as a possible statistic was not always understood. Even those candidates who produced a correct sampling distribution for R did not normally use their answer to (b) but started again with the samples and recalculated all the probabilities. Some candidates were confused about the range and it was not unusual to see values of 6 and 12 given. A number of candidates who had correctly completed part (b) made no attempt at this part.

Question 3

This question was accessible in the earlier parts but parts (c) and (d) were more discriminating. Fewer candidates than expected scored full marks in part (a) as many did not fully specify the distribution often just writing down $f(d) = \frac{1}{5}$ neglecting to include limits.
Candidates could have used question 7 as a clue as to what was required.

Part (b) was very well done. Almost all candidates earned both the available marks. The standard method, using the formula given in the Formulae Booklet, was by far the most common method. A minority of candidates used the alternative approach using integration. Errors were rare: a few candidates found the variance only, rather than standard deviation.

Parts (c) and (d) were more challenging, but correct answers were seen. The most common incorrect answer to part (c) was 0.5 as many candidates thought the question required the probability that it was less than the true weight. In part (d) many candidates mistook within 1 gram of the true weight to mean between 0 and 1 rather than between –1 and 1.

The majority of candidates realised that a Binomial distribution was appropriate in part (e). The marks scheme allowed credit for use of a Binomial distribution consistent with the answer in part (d), so that most candidates were able to earn at least two marks here.

Question 4

On the whole, question 4 was the most accessible question on the paper with 30% of candidates scoring full marks. It was very common to see all four marks being scored for part (a). Many different methods were used to solve the simultaneous equations. There is a standard method for this admittedly rare type of simultaneous equations: to divide the variance equation by the mean equation to obtain $1 - p = 0.85$. This technique was indeed seen, but not frequently. Another efficient method, using 'substitution', was to replace *np* in the variance equation by 4.2. Other techniques were used, mostly successfully, but often took several lines of working.

Fully correct solutions to part (b) were seen, but these were not widespread. Many miscalculated the expectation by using their values for the green sweets given in part (a) rather than the values given to them for the red sweets in this part of the question. Others incorrectly saw the need to round and attempted $\mathrm{E}(X) = 8.75 \approx 9 \quad P(X > 9) = \mathrm{P}(X \geq 10) = 1 - \mathrm{P}(X \leq 9)$.

The majority of candidates used the standard approach to the hypothesis test in part (c) by setting up hypotheses using correct notation and using the appropriate test statistic to calculate $\mathrm{P}(X \leq 1)$. There were, however, a number of common errors including calculating $\mathrm{P}(X < 1)$ or $\mathrm{P}(X > 1)$. Some candidates used the unnecessary approximation $\mathrm{Po}(4)$ distribution. The vast majority of conclusions were given in context, but there were still a few candidates who believed that the test showed that the shop's claim was not supported.

Question 5

This question was the most challenging on the paper and only 5% of candidates successfully scored full marks. Part (a) was accessible and most candidates scored full marks. Where there were mistakes it was usually in dealing with the inequalities e.g. $P(X > 12) = 1 – P(X \leq 11)$. Part (b) was also generally well done although a few candidates followed correct working with the conclusion that $k = 11$ rather than $k = 10$. Again some candidates found the inequalities difficult to cope with, particularly when a probability statement had to be reversed. Frequently $P(X > k) = 1 – P(X \leq k – 1)$ was seen.

In part (c), most candidates correctly worked with Po(5). The common error was in failing to square the $P(W = 4)$. Candidates generally used the formula to calculate $P(W = 4)$ rather than find $P(W \leq 4) – P(W \leq 3)$ from tables.

Part (d) of this question was clearly the most demanding part of the entire paper and only the most able were able to successfully tackle it. It was rare to see fully correct solutions as many failed to realise that this was a conditional probability and took no account of the information that the total for 2 months was 21. It was common to see $[P(X \geq 10)]^2$ given as the answer. A few managed to pay some attention to the fact that there had been 21 lost packages in two months and using Po(10) found $2 \times P(X=10) \times P(X=11)$. Of those who managed to recognise that this was a conditional probability many frequently assumed independence and over simplified the question.

Despite the difficulty in part (d), many persevered and went on to gain full marks in part (e). This part was often fully correct and most candidates were able to identify the correct approximation and cope with the continuity correction and the standardisation. A few were confused by the “greater than” probability and the negative z value so incorrectly subtracted their 0.976 from 1.

Question 6

The modal score for part (a) was five out of six. The majority of candidates used a z-value of 1.64 or 1.65 rather than using the more accurate value from the tables or from a calculator. Attention must be drawn to the instructions on page 1: ”Values from statistical tables should be quoted in full”. Here 1.6449 was required for full marks to be scored in part (a). Other details were sometimes incorrect. N(40, 40) or N(42, 42) were seen from time to time. A few candidates had difficulty manipulating inequalities, their last line being “$n < 55.7$ so $n = 55$”.

Part (b) seemed to cause some difficulty. An incorrect distribution was often seen. Some candidates provided two tails. Other candidates did in fact state only one tail, but the left-hand one. The notation for the critical region was generally sound. Very few ‘probability statements’, eg $P(X \geq 15)$ were seen. A small number of scripts featured what appeared to be a critical value rather than a critical region, i.e. $X= 15$.

Question 7

Overall this question provided a good source of marks for many candidates. Though full marks were only scored by the most able with around 15% doing so. Part (a) was generally well done with a clear layout and sufficient working shown to establish the proof. Very few attempted to verify rather than prove. The algebra was simple and few errors were seen at this stage. In part (b), very few candidates completed the required sketch accurately. Many candidates drew a quadratic segment with the correct curvature followed by a straight line segment with negative gradient and the salient points on the x axis were labelled. The common errors included joining the quadratic segment to the straight line or starting the quadratic segment from the x axis. Some candidates needed several attempts and some made no attempt at all.

Even if the graph had been drawn incorrectly most graphs had the correct mode at $X = 3$. However a few candidates still gave f(3) as the mode and not 3 in part (c). Part (d) was well answered by the majority of candidates. Most knew to work out $E(X^2)$ and then to subtract $[E(X)]^2$ and a good proportion of them got the correct answer. There were quite a number of opportunities for mistakes but the standard of working here was quite high. Slips included finding $E(X^2) – [E(X)]$ or errors with integration.

In part (e) the easiest method of using F(1) = 0 to evaluate c and F(4) = 1 to evaluate d which enabled independent calculations was missed by many candidates who instead integrated f(x) from scratch. Many of them went on to get correct values for c but lost accuracy finding d forgetting to add F(3) to the answer obtained from integrating the 2nd part of f(x).

Part (f) provided an easy end to the paper for many candidates. Almost all equated 0.75 to the appropriate line of F(x) although a few used 0.25 and a very few substituted $x = 0.75$ in the expressions. Most used the 2nd line of the expression for F(X) although there were some who used the 1st part and some used both. Having solved the resulting quadratic most then rejected the solution which was out of the range.

Pearson Education Limited. Registered company number 872828
with its registered office at 80 Strand, London, WC2R 0RL, United Kingdom