Examiners' Report

Principal Examiner Feedback

Summer 2017

Pearson Edexcel International A-Level in Mathematics Statistics (WST01)

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 2017
Publications Code WST01_01_1706_ER

IAL Mathematics Unit Statistics 1

Specification WST01/01

General Introduction

Overall there was mixed performance on this paper from students. Though many did display good statistical knowledge throughout, quite a large number of students were clearly not prepared for some topics on this specification, notably the normal distribution and discrete random variables. Question 1 parts (a) – (c) and question 5 parts (a) – (f) were the most accessible. At the top end, questions 1(d), 3(c), and 5(h) discriminated the most able students. Parts (e) – (g) of question 6 were the least successfully answered parts of the entire paper with even the most able students finding these parts challenging.

Report on individual questions

Question 1

Generally this question provided a good introduction to the paper and was a good source of marks for most students. Part (a) was generally well answered but there remain a large number of students whose errors inevitably involved getting the class-width and/or class boundary incorrect. It was also not uncommon to see rounding errors leading to an answer of 61.7

In part (b), nearly all students were able to correctly find the mean of the frequency distribution with any errors tending to be slips. Part (c) was also generally done well with the usual main errors here being to forget to square root the variance, not squaring their mean and occasionally dividing their mean by 50.

Part (d) proved difficult for most students, with very few realising linear interpolation was required. Of those who did, the result was often inaccurate due to oversimplifying the question by assuming 70 was the class midpoint. The majority assumed a normal distribution, with their previously calculated mean and standard deviation thus scoring no marks.

Question 2

Many students surprisingly found the work with box plots and outliers challenging and fully correct responses to parts (a) and (f) were rare. In part (a) most students ignored the outlier when finding the range. There were also those that wrote down an interval instead of a value. Finding the IQR in part (b) was much more successful as was using quartiles to determine the skewness of the distribution in part (c).

There was little trouble calculating the correlation coefficient in part (d), but written expression often lets students down when justifying its use, as was the case in part (e). 'Far away from 1' was the most common insufficient description of the lack of correlation between house prices and distance from work. Some students argued that as the r value was positive, the belief was not supported, but this alone would ignore that the magnitude is very small, in which case the sign is insignificant.

In part (f), many students ignored the instruction to show their calculations despite being told the rule for determining outliers. It was disappointing to see such a large number of students who were unable to draw a standard boxplot, this despite one already appearing at the start of the question. Many, of those who did not show their calculations, thought that the changed times would affect the quartiles.

Question 3

Students who draw sketches tend to perform better on normal distribution questions and this was indeed the case here. Part (a) was usually very well answered. For those who only scored partial marks, the two main errors were truncating and using $z = 1.3$ instead of the more accurate $z = 1.33$ and also not realising that subtraction from one was required. This would have been more obvious to students if they had drawn a sketch.

There was a good standard of responses seen in part (b), although some students lost the final accuracy mark because they did not express their answer correct to the four significant figures that the question demanded. A loss of accuracy was also noted by those who used a rounded standardised z-value of ±0.25 instead of the value 0.2533 which should be obtained from the percentage points table. A quite common misunderstanding was to use a probability value, particularly 0.4 as standardised z-value to solve for the minimum distance.

Many did not attempt part (c) of the question. Often solutions tended to be jumbled as students struggled with realising that conditional probability was a key concept. A probability of 0.5 leading to a z-value of 0 was used regularly here. Of those who did produce a suitable solution, some lost the final mark by truncating their final answer to 3.8 instead of using the required 3 significant figures.

In part (d), a large number of students thought the only requirement here was for a jump of greater than 4.1 metres, not realising that it needed a student to firstly qualify for having a second jump. This meant that there were many final answers of 0.0918 with no attempt to multiply by 0.4 whilst others calculated 0.0918^2. Not many realised they could use their answer from part (a) and hence started again to calculate $P(X > 4.1)$.

Question 4

There were many fully correct responses seen here, though, equally, many students scored 0 marks on this question. Those who realised that they needed an equation consisting of two terms in p in part (a) usually went on to score full marks for this part. A common incorrect attempt was to see the simple equation $0.4p = 0.26$, giving $p = 0.65$

Part (b) was more challenging but, generally, those were successful in part (a) tended to have success here as well. Even those who scored zero in part (a) for the answer of 0.65 often went on to score the first three marks here by setting up and solving an equation in the correct form.

In part (c), many just used the lower part of the tree diagram with a solution of $(1 – 0.15 – q)$ often seen as they incorrectly assumed that because there was no branch for cycling in the upper part of the diagram it meant that there was no need to involve that part in this calculation. Those who did bring in the upper half of the diagram usually scored the method mark even when they had incorrect answers for p and q.

Most students attempted a conditional probability in part (d). Quite a few scored either or both of the method marks by having either their p value as the numerator and/or (1 – their answer to part (c)) as the denominator. It is always disappointing to see such conditional probability expressions where the numerator is greater than the denominator and this was often seen here. Some students did not realise that they only needed to subtract their answer to (c) from one to find the denominator and instead they went back to the tree diagram and calculated the products of the four branch endings that resulted in 'not cycling'. Similarly many numerators were seen as $(0.44 \times 0.6) + (0.44 \times 0.4)$, with students not realising that it was simply p that was needed here.

Question 5

This was one of the more accessible questions on the paper, but part (h) proved most discriminating. There was no trouble calculating the required value in part (a) and most students were able to sufficiently explain that since r was close to 1 it supported the linear regression model. Part (c) was not answered well with many leaving it blank. Others tried to suggest that hours of sunshine can be fixed or controlled, clearly not appropriate in this context. More common was an interpretation of correlation 'as hours of sunshine increases, midday temperature increases', gaining no credit.

Part (d) caused some trouble for students who attempted to calculate the value directly rather than working backwards from the given correlation coefficient. In part (e), calculating the regression line equation was very well-practised and it was pleasing to see virtually all students give it in a suitable form and to a suitable accuracy without it actually being specified in the question. There were a significant number of students who mixed up the variables leading to an equation with temperature as the explanatory variable. Part (f) caused little difficulty for those with a correct equation in part (e).

Most students used the regular formula to calculate the standard deviation in part (g) with a good deal of success. However, in part (h), most did not make any use of this standard deviation to calculate the range of hours of sunshine. Many simply just wrote down 'reliable since interpolation' thinking that since all values of s were within 2 standard deviations of the mean, that this meant the 5 was also in range.

Question 6

This was, on the whole, the most challenging question on the paper. Students did not have sufficient practice with discrete random variables in context. Here the context was crucial to understanding the probabilities required. Despite the guidance offered in the earlier parts of the questions, there was very limited success in parts (e) – (g).

In part (a), the majority of students were able to show the given answer, although a minority either left part (a) blank or fiddled their answer using $\frac{0.432}{3} = 0.144$ or $0.6 \times 0.24 = 0.144$.

Nearly all answered part (b)(i) correctly, however, part (b)(ii) was not well answered as the vast majority ignored the fact that the sum of the probabilities should equal 1. Most arrived at the wrong answer 0.0864 (forgetting the 4-tail possibility) and followed this through in the rest of the question.

Nearly all students produced a correct expression for the expected value in part (c) using their probabilities. The success rate was almost as high in part (d) though some students still forget to square the mean when finding the variance, whilst others stop after only finding $E(X^2)$.

Part (e) onwards required a complete understanding of the experiment and many students struggled. Fundamentally, most students did not really understand the experiment being carried out as indicated in the poor level of response in part (e). The most common incorrect response being ‘a coin has two sides, heads and tails, so the number of heads is either 0 or 1’ followed by a probability distribution with probabilities of 0.5 and 0.5 or 0.6 and 0.4, for 0 and 1 respectively. Very few could explain that the experiment stopped once a head was obtained and would stop anyway at 4 spins even if no head had been obtained.

There was virtually no understanding by students about the relationship between the random variables H and X so any marks scored in parts (f) tended to come by luck.

Finally, part (g) was not very well attempted. Many missed out this part completely. Of those that attempted it, the majority had a $P(S = 1) \neq 0$ in their table and most did not list a value of $S = 5$ at all. Many had probabilities that totalled more than 1 in their table.

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