edexcel

Examiners' Report

January 2015

Pearson Edexcel International Advanced Level in Statistics Mathematics S1 (WST01/01)

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 2015
Publications Code IA040678

WST01/01 January 2015

Introduction

The paper proved to be accessible to all candidates with all questions having a gentle start but most having final parts that proved more challenging.

Comments on individual questions

Question 1

Most candidates answered parts (a) and (b) successfully. There were a few arithmetic errors but most could form equations for b or c and usually the correct values were found. Part (c) was generally answered well too although a number wrote down P($X <$ 4). Parts (d) and (e) were only answered correctly by the stronger candidates and then not always in the most efficient manner.

In part (d) many tried listing all 9 cases and invariably either forgot to include cases like 1, 3 and 3, 1 or they missed the "doubles" such as 1,1. Finding the probability that X is odd and squaring it was rarely seen. The major problem in part (e) was that candidates didn't seem to realize that a conditional probability was required and those that did sometimes missed one of the probabilities on the numerator such as one of the 1,5 pair or 3,3.

Question 2

This question was answered well with over half of the candidates scoring 7 or more. In part (a) a large number of candidates did not realize that the number of students in each year was different and many used 31 to calculate the median of Year 11 and arrived at an incorrect answer of 53 but the median for Year 7 was usually correct.

Part (b) caused few problems although 38 was often given instead of 42 for the upper quartile. It was disappointing to see some candidates having $Q_1 > Q_3$ and being perfectly happy to use these values in part (c).

In part (c) only a small minority confused the conditions for positive and negative skew and some did not use the quartiles but attempted to answer the question using the mean or the mode.

In part (d) many candidates knew that the skewness of the data was a reason for not modelling with a normal distribution but only a few mentioned that the data was discrete and therefore a continuous distribution was not appropriate. Some gave their second reason that the mean was not equal to the median but this is equivalent to stating that the data was skew and could not score the second mark.

Question 3

The first 3 parts were answered very well and two thirds of the candidates scored 6 or more on this question. Most showed their working clearly in parts (b) and (c) and were able to pick up some marks even if an earlier part was incorrect. Disappointingly a sizeable minority thought $\sum bm$ meant $\sum b \times \sum m$ and a number failed to score the accuracy mark in (c) because they only gave their answer to 2sf and not 3sf or better.

Part (d) was understandably less well answered with many simply giving a description (positive correlation) and only a few gave a full interpretation referring to the price of milk and the price of bread. Part (e) was only answered fully by the best candidates. Many simply said that r was not affected by coding and some stated that the value of r will decrease but it was rare to see a convincing reason given and an understanding of what r is actually measuring is clearly not grasped by the majority of the candidates.

Question 4

Despite the algebraic nature of this question it was usually answered very well and nearly 50% of the candidates gained full marks. Parts (a) and (b) caused few problems although a popular misconception was that P(A) = x and P(B) = y and an answer of x – 0.1 was a common error in (a)(i).

In part (c) most attempted to form an equation using P(A) = 2P(B), although poor use of brackets meant they often could not simplify it correctly. A sizeable minority though failed to write down a second equation in x and y using the fact that the total probability = 1 and so they were unable to solve two linear equations and obtain values for x and y.

Question 5

There was a good spread of marks on this question with weaker candidates usually answering parts (c), (d) and (e) well but part (g) offering a challenge for the more able. In part (a) most chose h as the response variable but they could not always provide a satisfactory justification. The instruction to provide an "interpretation" in part (b) means that the examiners are looking for an answer that uses the context and gives appropriate values.

There were plenty of good answers explaining that for every extra minute of exercise the resting heart rate drops by 0.43 but most did not include all the relevant detail. Parts (c) and (d) were answered very well as was part (e) except that a few candidates failed to pay attention to the units and simply substituted t = 1 into the regression equation rather than t = 60. Many stated that the estimate was reliable because "it" was in the range. The key point is that the value of the explanatory variable t is within the range of values for t and responses that did not specify this gained no credit.

Part (g) caused some confusion and many got lost in a sea of equations involving a, b and 0.95. Those who did make some progress and standardized with a or b, and set their expression equal to a z value greater than 1, often used a z value of 1.64 rather than 1.96 and lost the accuracy marks.

Question 6

Nearly a quarter of the candidates gained full marks here but there was a good spread of marks with most making progress on the first 4 parts of the question.

Very few could not interpret the probability function correctly and most knew how to show the result in part (a) clearly by equating the sum of the probabilities to 1. Some seemed not to read or understand the $\neq$ symbol correctly and gave the answer for P(X = 4) in part (b).

Part (c) was answered very well although some did not give an exact answer. The usual problems occurred in part (b). Many can calculate E(X^2) correctly but some thought this was Var(X) and a few of those who realized there was more work to do only subtracted E(X) rather than $[\mathrm{E}(X)]^2$ even after quoting a correct formula. Others failed to score the final mark for failing to provide an exact answer. Part (e) proved quite challenging and many of the usual errors of confusing E(Y^2) with $[\mathrm{E}(Y)]^2$ were seen.

There were two commonly seen successful approaches: one involved forming the probability distribution for Y and effectively repeating the steps in (c) and (d). Some though did realize that $\mathrm{E}(Y^2) = \mathrm{Var}(Y) + [\mathrm{E}(Y)]^2$ and then used the formulae for E(aX + b) and Var(aX+b) to obtain E(Y) and Var(Y).

Question 7

The normal distribution remains a challenging topic for some and nearly a quarter were unable to gain more than a couple of marks.

In part (a) most could standardize and many had P($Z > -1.94$) but it was disappointing to see them giving their final answer as $1 - 0.9738$. Part (b) caused problems for many as they missed that P($W < k$) = 0.75. A common error was to standardize with $+k$ and $-k$ and then multiply one expression by 3 and form an equation for k. Those who did start correctly were often able to complete the problem but some failed to give their answer to 1 decimal place and others used a z value of 0.68 instead of 0.67. Candidates are not expected to interpolate in the tables but they should pick the value closest to the one required.

The answers to part (c) were strange and varied. Some, having found the lower quartile, gave this; others stated that k was the mean or the median and some even thought k stood for kitten! It was encouraging to see candidates who could not attempt part (b) making headway with part (d). Some failed to use the percentage points table and used a z value of 0.84 rather than 0.8416 (or better from their calculator) and a few fiddled their working to arrive at a positive value for σ. An equation of $\frac{116-120}{\sigma} = 0.8416$ leading to an answer of $\sigma = 4.75$ will not score the final accuracy mark.

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

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