Examiners' Report
Principal Examiner Feedback

January 2018

Pearson Edexcel International A Level
In Statistics 1 (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

January 2018
Publications Code WST01_1801_ER

IAL Mathematics Unit Statistics S1

Specification WST01/01

General introduction

Most of the paper was accessible to all the students and many made very good progress on questions 1 – 5. Questions 6 – 7 were notably more challenging and provided good discrimination at the top end. Students are generally much better now at carrying out standard statistical calculations but interpreting these, as was required in question 1(e) is still proving a challenge to most.

Report on Individual Questions

Question 1

It was rare to see an error in part (a) but many students did not know a suitable formula for finding the variance and failed to score marks in part (b) or part (d)(ii). Part (c) proved to be a good discriminator with many failing to see the connection with part (b) but others identifying class *B* because the variance was larger. Very few had any problem in finding the correct mean in part (d)(i) but good answers to part (e) were far less common with weaker students failing to give supporting reasons or simply stating that coding makes no difference. In part (i) many mentioned that the variance would not change because the data was coded which is of course insufficient. The key here is that variance is unchanged because the "coding" only involves addition. Part (ii) was more often correct and the examiners did condone the calculation of the new mean, as this was so straightforward, though all that was required was a statement that the total score had increased. Part (iii) was the most challenging: many simply stated that the variance would increase because the marks were only added to one class but, of course, had they been added to class *B* then the mean of class *B* would be closer to the mean of class *A* and the opposite conclusion would apply.

Question 2

Part (a) proved surprisingly challenging with many forgetting to include $A \cap B$ in their shading or shading $A \cap B'$. In part (b) most obtained $\mathrm{P}(C \cup D) = \frac{1}{2}$ but few stated $\mathrm{P}(C \cap D) = 0$ though some were credited with the mark if they gave a correct answer to P($C|D$). In part (c)(i) many quoted the correct addition rule formula and were able to substitute at least one value correctly though the incorrect form $\mathrm{P}(F \cup G) = \mathrm{P}(F) + \mathrm{P}(G)$ was often used. The correct rule for independence was often used too. One common error was to state that $\mathrm{P}(G) = \frac{3}{8} - \frac{1}{6} = \frac{5}{24}$. In (c)(ii) it was rare to see students simply stating the independence condition $\mathrm{P}(F \mid G') = \mathrm{P}(F)$ and writing down P(F) but many correct answers followed from some fairly complicated working.

Question 3

This question was answered well by most students. Responses to part (a) were often correct though some students did not mention the strength of the correlation but merely its sign. In part (b) many mentioned the dependent variable and some gave clear explanations in context. Parts (c) and (d) essentially required the students to calculate the equation of the regression line which many did successfully. The other two marks though required some careful explanations. In (c) many did not realise that a simple calculation based on the gradient ($100 \times b$) would have been sufficient to confirm the statement, some used the whole equation to try and answer this thus requiring the substitution of two values for c and c + 100 and then a clear subtraction to confirm the statement. In part (d) more were able to quote the intercept value to give the expected number of hours of sleep when no caffeine was consumed.

Question 4

Most students made some initial progress with this question but only the more able scored 10 or more marks here. A correct expression for E(X) was usually obtained in part (a) but the statement about Var(X) led many students to find a similar expression, in terms of a and b for E(X^2) and often these two linear equations (both expressions mysteriously set equal to 0) were solved. Those who did find a correct 2nd equation based on the sum of the probabilities equalling 1 usually assumed E(X) = 0 and went on to solve their equations successfully. The students should note that where a specific demand is made "write down the value of E(X)" we will expect to see an answer given for this and so a number of students lost the mark for (b)(i). Those who had genuine probabilities for a and b (there were plenty with negative values) could usually find Var(X) and multiply this by 9 to answer part (c). Many students had a stab at (d) and often a correct, or correct follow through, answer was seen for (i) though part (ii) proved more challenging.

Question 5

This was the most successfully answered question on this paper. Parts (a) and (b) were answered well and many knew that "coding does not affect correlation" in part (c) though a number of other students simply gave an interpretation of the value of the correlation coefficient. In part (d) many tried to substitute for w but there were frequently errors in simplifying their equation and the intercept was often incorrect. A few attempted to go back and find the regression equation from scratch but this was rarely successful. Part (e) was answered very well: some using x = 21 000 and others coding to get w = 1 and using this is the given equation. In part (f) we expected to see some calculations to either code the given salaries or convert the maximum and minimum values of w in the data to values of x. Many were able to do this but sometimes an incorrect conclusion was made. Strangely some coded the given salaries to get w = 5 and w = 20 but then compared these with the first and last values of w in the table and concluded that the estimates were outside the range of the data and would not therefore be reliable.

Question 6

This question was not answered well with nearly half (48%) of the students scoring zero. A small number misread the first line and worked with 7 socks (2 blue and 5 other colours) and the standard misread rule was applied in these cases.

Many clearly did not understand the situation properly as there were many non-zero answers offered for part (a). More made some progress in part (b) even if it was only to state $1-\mathrm{P}(S \leqslant 2)$. In part (c) one of the cases $B'BB$ or $BB'B$ were often missed and some included the BBB' case. Part (d) was the most challenging part and only a few realised that a conditional probability was required and even fewer gave a correct ratio. In part (e) a number successfully gave one correct product of probabilities but it was rare to see all 4 and the correct answer.

Question 7

The standard probability calculation involving the normal distribution in part (a) was answered quite well though it is disappointing to see some have a correct statement such as P($Z > -0.4$) followed by a probability less than 0.5 The rest of the question got progressively more challenging. In part (b) many muddled their way through to obtain $\frac{k-180}{15}$ = to a z value with $|z| > 1.5$ and many had a value of 1.96 but the correct sign was less frequently seen so only the stronger students had a correct answer of 150.6 Part (c) provided a good challenge for the best students at the end of this paper. There were plenty of attempts at standardising with the relevant means and standard deviations but few were able to form a correct equation in w, the usual problem was a missing minus sign which led to an incorrect value of $w = 144$ A further mark for attempting to use this value was often obtained but only the very best students achieved the correct answer of 0.212

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