Examiners' Report/
Principal Examiner Feedback

Summer 2016

Pearson Edexcel International A-Level
Statistics 1
(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

Summer 2016
Publications Code WST01_01_1606_ER

Statistics S1 (WST01)

General introduction

The paper was accessible to all the candidates and most made very good progress on questions 1 – 4. Questions 5 – 7 were notably more challenging and provided good discrimination at the top end. It was pleasing to see that work on the normal distribution is improving. Determining all possible outcomes to calculate a probability remains challenging for most candidates. Working confidently with conditional probabilities continues to be a topic only accessible by the most able candidates. As always, candidates should use exact values in calculations to avoid losing accuracy in their final answers.

Report on Individual Questions

Question 1

Candidates performed well on the opening question of the paper with 50% scoring at least 11 out of the 12 marks here. Part (a) was answered well by the vast majority of candidates. The most common mistake was a slip in finding S_{wp} and candidates should be reminded to check over their answers particularly when they are used in later parts of the question. The main issue with part (b) concerned truncating rather than rounding as –0.929 was a common incorrect answer.

A large number of candidates failed to contextualise their answer in part (c) merely stating 'negative correlation' or only used the letters w and p rather than *weight* and *percentage of oil content*. Part (d) started well for most candidates who worked out a correct expression for b. Rounding the value of b before finding a led to a loss of accuracy. Even when using a more accurate value for b some still gave 2 significant figure coefficients in their final regression equation thus losing the final mark. It was pleasing to see that few candidates attempted the regression equation of w on p and even fewer gave an equation in terms of y and x. It is important to note that exact fractions are not accepted in the final regression equation.

In part (e), most candidates correctly substituted $w = 60$ into their regression equation and many gained both marks even with slightly inaccurate regression equations

Question 2

Two thirds of candidates scored full marks here whilst a quarter were unable to make any progress at all. Those candidates who answered this question with confidence were those who were able to calculate that 1 student was represented by an area of 0.9cm^2. Some candidates understood how to calculate the area of the bar but were then unable to relate it to frequency of 20. Others were able to establish the fraction 20/18 but then made the mistake of dividing by this value rather than multiplying which gave a common incorrect answer of 85. Some were unable to distinguish between which values were areas and which were frequencies, leading to obviously wrong answers including values of less than 20. There were a few candidates who worked with 1.11 rather than 20/18 and failed to round 104.895 to 105 thus losing the accuracy mark.

Question 3

This was the second most successful question on the paper with 30% of candidates scoring full marks. Despite this, candidates often missed out part (a) and it was disappointing that so few were able to put a name to a basic distribution. The word ‘uniform’ was often missed out and some gave incorrect responses such as normal distribution, random variable or discrete distribution.

Parts (b) and (c) were generally well answered with many candidates recognising that F(3) was a cumulative probability.

In part (d), many candidates were able to simplify the inequality but in some cases they did not use any probability notation and once $X > 3.5$ was obtained no further working was seen. Some successfully constructed tables to find $3X – 3$ for all X (1 to 5) and $X + 4$ for all X (1 to 5) and compared these values to see where the inequality was true.

Parts (e), (f) and (g) are standard for most candidates and as such very few mistakes were seen. The incorrect formula $\text{Var}(X) = \text{E}(X^2) – \text{E}(X)$ still makes an appearance in a significant minority of solutions. Very few candidates used the standard formulae for the discrete uniform distribution and instead completed these calculations in full.

Finally, part (h) was more challenging for candidates. Many were in a rush to show knowledge that $\text{Var}(aX – 3) = a^2\text{Var}(X)$ but never attempted to find the value of a. Those who successfully found the value of a sometimes spoiled a good response by subtracting the 3 from $a^2\text{Var}(X)$.

Question 4

Whilst all candidates could access the first parts of this question, part (f) was a clear discriminator of the most able. Virtually all candidates answered part (a) correctly with only a few errors in finding b. In part (b) the mean was generally calculated correctly although a significant number of candidates wasted time re-calculating the given $\sum fm$ – some incorrectly. One worryingly common error was to divide by 5. Although the majority of candidates were able to calculate the standard deviation in part (c) correctly, there were many responses where only the variance was calculated, $\sum fm^2$ was not divided by n, or the wrong value of n (often $n = 5$) was used. Accuracy marks were also regularly lost through the use of a rounded value for the mean.

There has been a notable improvement in the use of interpolation to find the median with the majority of candidates in part (d) able to identify the correct fraction and class boundary. Where errors occurred, candidates usually misinterpreted the class boundaries for the specified class as 4.5 to 10.5. In part (e) most correctly identified mean > median and positive skew though many unnecessarily used 3(mean – median)/standard deviation. Attempts to use $Q_3 – Q_2 > Q_2 – Q_1$ were generally less successful as the quartiles were often not correctly calculated or not attempted at all.

Dealing with the coding of the data highlighted varying degrees of competence and understanding. Fully correct responses in part (f) were not common and there were many blank responses to this part. Of those who did attempt part (f) the mean was generally decoded successfully, although some were let down by poor algebraic skills, adding 5 before multiplying by 10. The decoding of the standard deviation was far less successful. Far too many candidates incorrectly added the 5 whilst others claimed that the standard deviation is not affected by coding.

Question 5

This was the least successfully answered question on the paper as 50% of candidates failed to earn more than 1 out of the 8 marks here. Many candidates were unable to identify the correct probabilities from the information in the question. Those who attempted to list all of the ways of scoring 2 were able to make good progress in part (a). A very common error was not taking into account P(2,0) and P(0,2). Others calculated P(1,1) twice. Candidates need to understand when the order is important in probability questions.

Again, incomplete responses were seen in part (b) as candidates could not precisely identify which outcomes, and hence which probabilities, were required. Many misunderstood the question, and used 3 and then 0 as a probability, rather than realising that once a 3 was rolled, the die was not rolled again.

Candidates found part (c) particularly difficult and were inclined to ignore the conditional probability altogether simply subtracting ½ from their answer to part (b). Candidates need to be aware that the term ‘Given that...’ implies conditional probability and therefore a ratio of probabilities is required.

Question 6

This was a challenging Venn diagram question that required both strong algebraic skills and a good understanding of statistically independent events. The majority of candidates gained the mark in part (a) but some confused the union and intersection sign, stating the answer as 0 because the two events are mutually exclusive.

A fully correct solution to part (b) was not often seen. Many applied the addition rule correctly but struggled to show understanding of independence. There were a large number of attempts which relied on circular reasoning, i.e using $\mathrm{P}(B) = \frac{3}{8}$ to show $\mathrm{P}(A \cap B) = \frac{3}{20}$ to then show that $\mathrm{P}(B) = \frac{3}{8}$. Some had fully correct substitution into the addition rule, but were unable to manipulate the algebra to show the given answer.

It was rare to see concise answers to part (c) which used the fact that *A* and *B* are independent. Most adopted the longer method, using the conditional probability formula but many correct answers were still eventually found.

There were two types of Venn diagram that could be accepted in part (d). The more popular one was where three circles were drawn such that *B* intersected both *A* and *C* but *A* and *C* did not intersect. One quite common error in this diagram was the circle for *C* seen separately, not intersecting either of the other two circles. The other style of diagram was where there were three circles intersecting, but candidates should be aware that blank spaces on the diagrams are not considered 0 and these must be explicitly labelled. Most candidates successfully achieved the first method mark for the correct probabilities in the regions $(A \cap B \cap C')$ and $(A \cap B' \cap C')$ but very often the next mark was lost because there wasn’t a correct probability in the outer region of the diagram. For the next mark, many realised that the sum of the regions in *B* had to add up to $\frac{3}{8}$ and those in C to $\frac{1}{2}$. Overall in this part, scores of either 2 or 3 marks were the most common.

Question 7

As expected, candidates found this question on the normal distribution challenging but less able candidates did make a good attempt at some parts. Overall, this was the second most difficult question on the paper. In part (a)(i), most candidates now seem to know how to standardise with very few using 1.6^2 as the standard deviation. There are still a few who did not appreciate that their resulting probability of 0.8944 needed to be subtracted from 1 and these tended to be the ones who did not draw a diagram. In (a)(ii) many correct answers were seen but most did not take into account symmetry here and made things far too complicated.

Part (b) proved to be the most demanding part of the entire paper and it was rare for candidates to see the link between the 1006 and the mean. Very few were able to use the given information to find the area of one of the tails and many simply used 0.9426 to obtain the common incorrect z-value, $z = 1.57$.

Finally in part (c), most candidates realised that simultaneous equations were required for this part, but had some difficulty in obtaining two fully correct equations. Candidates must use the second normal table (whenever possible) to obtain a more accurate (4dp or better) z-value. A common error was using +2.32(63) (incompatible signs) when standardising r leading to a negative value of q. This mistake should have been spotted by candidates. Again a diagram could have helped to avoid this. Despite many candidates making errors with incompatible signs and inaccurate z-values, the most able were able to work their way through this part to successfully obtain accurate values of r and q.

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