Examiners' Report/
Principal Examiner Feedback

January 2016

Pearson Edexcel International A Level
in Statistics (WST02)
Paper 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 2016
Publications Code IA043348

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

Statistics S2 (WST02)

General introduction

Overall, students were well prepared for this paper and, in most cases, were able to attempt all questions. The beginning of the paper (questions 1–4) was accessible to students at all levels and provided a good source of marks for many. Parts 5(c) and 7(a) discriminated at the top end. Questions which require an interpretation of the data given remain challenging for many.

Question 1

Though the majority of students found question 1 to be an accessible start to the paper, part (a) proved to be challenging for most. There were many definitions of sampling frames but most failed to give a reason as to why this database would not make a suitable one. Many students thought that using a database would be time consuming. The database not containing all of the customers was the most common correct answer given.

Part (b) was far more successful with the majority of students being able to correctly identify a binomial distribution together with the correct parameters. Part (c) was a routine calculation and therefore was a source of easy marks for most. A few students had $1 - P(X \leq 9)$ rather than
$1 - P(X \leq 10)$.

Question 2

In part (a), virtually all students were able to use $E(X) = 4$ to obtain $a + b = 8$, but a small minority were unable to establish a second equation from the given probability. Solving the two simultaneous equations presented no difficulties to the majority of students. Part (b) was successfully answered by a large majority of candidates. Others mistakenly substituted $E(X) = 1/8$ rather than 4.

The value of $E(X^2)$ was usually obtained in an exact form in part (c). A small percentage of candidates made a sign error in rearranging $\text{Var}(X) = E(X^2) - [E(X)]^2$ or used the incorrect formula
$\text{Var}(X) = E(X^2) - E(X)$. In part (d), most students were able to deal with the inequality and score both marks. A handful of students made slips in rearranging the inequality of stated $X > 4$ rather than giving a probability.

Question 3

Students had a high success rate with this question with just over half of all students scoring at least 9 marks here. A majority of students correctly identified the Binomial distribution B(60, 0.1) and followed the instructions in the question accurately. Some, however, rushed into an approximation before being asked whilst others gave an expression for $P(Y = 1)$ rather than $P(Y \leq 1)$.

Part (b) was better answered overall, as students were told to use a Poisson approximation in the question. It is pleasing to see many students correctly applying the Normal approximation with continuity correction in part (c). Many fully correct solutions were seen. Where students were less successful, the most common error was to use N(6, 6) as an approximation to the Po(6) approximation, rather than N(6, 5.4), the appropriate approximation to the original B(60, 0.1). It is encouraging that other potential errors such as forgetting to use a continuity correction or standardising with variance instead of standard deviation were less frequently seen on this paper.

Students had a number of correct reasons to select from and, overall, the response to part (d) was good. Stating 'n is large and p is small' was the most popular. Answers listing the assumptions required for a Poisson model rather than referring to the Poisson approximation were quite common and generally failed to score the mark.

Question 4

This question also provided a good source of marks for students with roughly half scoring full marks here. Most students correctly set up $F(2) = 1$ and went on to successfully show that $d = 2$ in part (a). Encouragingly, those who found ± 2 virtually always went on to reject $d = -2$ as a solution.

In part (b), $x = 1.5$ was usually substituted into the correct line of the distribution and a final answer of 29/64 was generally obtained. A minority of students incorrectly used $P(X < 1.5) = 1 - F(1.5)$. The lower quartile was correctly given as 1 by the majority of students, though some used the 3rd line of $F(x)$ to obtain the result when the 2nd line was clearly more efficient. Incorrect answers were 0.25 and, less commonly, 0.0625.

A common error in part (d) was to add both (2nd and 3rd) lines of $F(x)$ together before equating to 0.5, however most responses seen here were usually fully correct. Part (e) proved to be more challenging, with a significant proportion of students stopping after calculating $P(X > 1.9)$. Some then went on to mistakenly equate this back to the 3rd line of the cumulative distribution function. A minority either prematurely rounded or truncated 0.148395 to produce an inaccurate value for k as 0.592.

Question 5
At this stage of the paper, students were more challenged and question 5 produced a near uniform distribution of marks. Part (a) proved to be challenging for a large number of students. Those who had thought carefully before putting starting to answer produced efficient and accurate solutions here. Many students made the mistake of joining the two 10 year periods together. Less able students gave an answer of $1 - \mathrm{P}(X = 0)$ or multiplied this by 2 instead of squaring it.

Part (b) was far more successfully completed and, for some, proved to be the only 2 marks gained in the entire question. Students persevered with part (c), even though the demand of this question caused problems for some. Those who noticed the useful hint in the question to 'Use the tables ...' were generally able to make a successful start by writing $\mathrm{P}(X = 4) = \mathrm{P}(X \leq 4) - \mathrm{P}(X \leq 3)$. Some found that for $\lambda = 1.5$, the probability was close to that required and gave a common incorrect answer of 15 years. Others having identified Po(8.5) as the correct distribution gave 8.5 as their final answer.

Most students are aware of the necessity to express the null hypothesis for a test of a Poisson parameter in terms of λ (or μ) and part (d) was answered correctly. Some ignored the demand in this part and gave both hypotheses.

Finally, in part (e), a pleasing number of students correctly identified the critical value. It is still common to see responses stating that the critical region can be expressed as $c = 16$ or $\mathrm{P}(X \geq 16)$.
A surprisingly common error was to incorrectly use a two-tailed test/critical region here.

Question 6
The most able students were able to demonstrate their ability here with a quarter of all students achieving full marks on this question. There was a pleasing response to part (a), with nearly all students picking up at least the first two marks for attempting the correct integral here. Minor slips (with brackets or negative signs) meant some students fiddled their response but still arrived at the correct solution.

$b = -1/90$ was nearly always obtained in part (b), with a large proportion of students going on to show that $\mathrm{E}(X) = 5.4$. A few had approximated their fractions in decimal form and were unable to gain the final accuracy mark. A small minority of students did not substitute in the values of a and b and gave $\mathrm{E}(X) = 600a + 114b$.

Students invariably defined the distribution function for all values of x in part (c). Whilst many were completely correct, a significant minority either neglected to insert a constant of integration or used 0 instead of 1 for the lower limit of their integral. The main error in defining the intervals occurred on the first line of $\mathrm{F}(x)$, where $x < 0$ was occasionally seen.

Part (d) was received well by students with a significant majority being awarded both marks. There were, however, some students who opted to calculate $\mathrm{F}(\mathrm{E}(X))$ and lost both marks as a result. The final part of the question produced a more varied response. Most followed the instruction to 'use your answer to part (d)', but a small proportion compared their value of $\mathrm{E}(X)$ with the mode. Some students, having obtained a correct answer in (d) and having compared it correctly with 0.5, went on to incorrectly conclude positive skew.

Question 7

The final question on the paper was the most discriminating on the whole paper. Many students struggled with the unstructured nature of part (a) and the fact that it required multiple steps to arrive at the solution. Though attempts were made to find the probability that the fisherman catches at least 6 fish on one of the trips, these often stopped there. It was not uncommon to see Po(12) or N(12, 12) attempted here as students did not know how to use the information ‘exactly 3 trips’.

Part (b) of this question had mixed success. Those who correctly set up a two-tailed alternative hypothesis generally achieved well on this part. A very common error was to state $\lambda > 8$ as the alternate hypothesis. The majority of students were able to use $1 - P(X \leq 13)$ to obtain a correct probability for the test statistic. Reaching a correct conclusion at this point required students to compare the probability with 0.025 as this was a two-tailed test and this caused confusion for some. For those persevering to the end, it was pleasing to see correctly stated conclusions in the context of the question.

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