Pearson Edexcel

Examiners' Report
Principal Examiner Feedback

January2020

Pearson Edexcel International GCE
in Statistics Mathematics S2 (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

Grade Boundaries

Grade boundaries for all papers can be found on the website at:
https://qualifications.pearson.com/en/support/support-topics/results-certification/grade-boundaries.html

January 2020
Publications Code WST02_01_2001_ER*

General

The paper was accessible to all candidates and overall there were some strong performances. Each question proved to be accessible with later parts challenging the most able candidates. Qu 2(c) and Qu. 5(d) were found to be the most discriminating on the paper. The quality of the sketches of the probability density function in Qu 4(b) was generally poor. It was pleasing to see that the conclusions to the hypothesis tests were generally given in context.

Report on Individual Questions

Question 1

Part (a) proved to be a good start to the paper with most candidates earning both marks. Both methods, using the cumulative probability tables and the formula for Poisson probability, were seen frequently. The most common error was to use $\mathrm{P}(H \le 6)$ on its own. Some candidates made an error by writing down $\frac{e^{-4}6^4}{6!}$ rather than $\frac{e^{-4}4^6}{6!}$ thus losing marks.
Part (b) was also a good source of marks for a majority of candidates. Incorrect methods, however, were seen surprisingly often: $\mathrm{P}(J \le 7) - \mathrm{P}(J \le 3)$ and even $\mathrm{P}(J \le 6) - \mathrm{P}(J \le 3)$. Some tried to square the probability resulting from a Po(4) distribution rather than using Po(8).

A substantial majority of candidates scored all five marks in part (c). Marks lost in this part were mainly due to using a 29.5 instead of 30.5 or no continuity correction at all. On some occasions candidates forgot to use "1 – ...".

Part (d)(i) caused the most problems for this question. Irrelevant statements concerning the mean, np and λ were frequently seen. Some candidates just wrote 'it's large' without specifying that the probability is large.

In part (d)(ii), a few candidates failed to implement the helpful advice printed in the question relating to the bold "**not**". A majority of the remainder scored the three marks available, although a few scripts were seen where the incorrect probability $\mathrm{P}(L \le 5)$ was used.

Question 2

Part (a) was a routine question testing Binomial probabilities. Full marks were scored by a substantial majority. Most candidates worked with the cumulative probability tables for both parts, with only a minority using the formula for Binomial probability in (i).

Many confident solutions to part (b) were seen. Candidates adopted a clear strategy for the hypothesis test which was then implemented accurately. It was notable that a large proportion of candidates finished with a completely correct conclusion in context.

Understanding the context of part (c) made this part one of the most discriminating on the entire paper. Of those who made some progress, most realised the necessity of calculating the new probability of more large eggs than small, but some simply used the value of 0.45. Many candidates seemed not to understand that what was required statistically was the calculation of expected values, and were confused by the non-standard context. A common error was to subtract the cost of the supplement twice. Those who found two expected values then usually came to the correct conclusion.

Question 3

Once correct expressions were obtained in this question, candidates made good progress and showed good manipulation of algebra. Nearly half of candidates achieved full marks on question 3.

Most candidates earned the single mark in part (a), but some took many stages of working to get there.

Again, many fully successful answers to part (b) were seen. Inevitably, errors were seen in algebraic manipulation, but these were generally rare. Candidates are reminded to make their methods clear so that method marks can be scored when answers are incorrect. Although the most common method used by candidates was to use $\mathrm{E}(T^2) = \mathrm{Var}(T) + [\mathrm{E}(T)]^2$, those opting for an integration attempt, were equally successful. The latter method required more attention to detail with the algebraic manipulation. A continued common error is those candidates who omit the square on the $\mathrm{E}(T)$ term.

The overall response to part (c) was excellent, with a high proportion of candidates scoring full marks. The only minor error seen was those using $1 - \mathrm{P}(X \le 20)$.

Question 4

Part (d) was the most discriminating part of this question, but overall a good standard of responses were seen here.

In part (a) most candidates were not only familiar with the relevant theory and techniques, they also responded appropriately to the 'show that' instruction by giving a clear and fully detailed solution which lead to the required answer. Most integrated both parts of the density function, but a small number recognised that the function corresponding $1 \le t < 2$ was a rectangle.

The sketches in part (b) were disappointing. Very few candidates obtained the correct shape for both parts of the probability density function. Many candidates drew two straight lines. Some candidates were familiar with the correct general shape for the curved part, but they assumed that the density function started at its maximum. Other candidates realised that the right-hand end point of the straight line had a different vertical co-ordinate to the left-hand end point of the curve, but joined these two points with a solid vertical line. It was rare for candidates to score both marks in part (b).

In part (c), a significant number of candidates wrote down an incorrect answer of 2 or 3 without showing any working or offering any explanation. Some candidates appeared to be basing their conclusion on an incorrect sketch in part (b). The mark allocation in this part indicates that some work will be required. There were, however, a substantial majority who knew the correct strategy, i.e. solve $\mathrm{f}'(t) = 0$. This almost always led to the correct answer. A few candidates went further than required by differentiating twice in order to confirm that their answer did in fact correspond to a maximum value of the probability density function.

Many strong attempts at part (d) were made, but only the most able candidates went on to complete the question with full accuracy including giving the entire answer in terms of *t*. A majority of candidates used definite integration for both regions. A large number, however, forgot to add F(2) for the part of the cumulative distribution function corresponding to $2 \le t < 4$. Even though this is a routine type of question, candidates still often neglect to deal with the second part of the probability density function. Some otherwise excellent responses lost the final mark due to mixed lettering in their final answer.

Many correct responses to part (e) were seen, mostly using the standard method 1 – F(3). Other candidates successfully used the alternative approach of starting again by integrating the density function between 3 and 4. Some candidates evaluated F(3) but failed to subtract it from 1. Candidates with an incorrect 3rd line to part (d) were able to score the first of the two marks for part (e), but only if their method was shown. It was not uncommon for candidates to write a final (incorrect) answer without showing the details of the substitution of 3 into their (incorrect) 3rd line, losing one mark in the process.

Question 5

This question was the most demanding question on the entire paper with just over 10% of candidates going on to achieve full marks.

A large number of correct critical regions were seen in part (a). However, this was not universal. Errors in detail sometimes resulted in $X \geq 8$ for the upper tail. More serious errors were conceptual: $X = 0, X = 9$ suggests confusion between a critical region and a critical value while $0 \leq X \leq 9$ and $1 \leq X \leq 8$ both suggest confusion between a critical region and an acceptance region. Many continue to give CRs as probability statements.

A majority of candidates were able to obtain the correct 'actual significance level' in part (b). Though again there seemed to be some confusion amongst weaker candidates as some went on to do 1 – their probability.

While many candidates answered part (c) correctly, elsewhere some serious misconceptions were apparent, for example: “6 not in CR: does not support Chris's claim”.

The most problematical part of the question was (d). There seemed to be a significant number of scripts with no attempt at (d). It was not uncommon for candidates to score only the first M1 for a correct probability inequality (although an equation was allowed). Some candidates could not identify a correct version of this probability statement, and rarely made any progress at all. The standard method required solution of the exponential inequality (or corresponding equation) $e^{-\frac{2n}{9}} < 0.1$, the preferred method being the use of logarithms leading to $-\lambda = -\frac{2n}{9} < -2.3025\ldots$ Some candidates believed only values from the ‘tables’ could be used and hence went on to round 2.3025... upwards, leading to $\lambda = 2.5$. The parameters of Poisson (and Binomial) distributions do not have to conform to the limited set of values used in standard tables. Furthermore, in this case tables are not even required, since no more probabilities need to be calculated. Despite all the problems that were witnessed, there were some excellent solutions using a correct approach, accurate calculations and a fully-detailed yet concise method.

Most candidates were familiar with the requirements of a standard significance test. The only serious problem in part (e) was a lack of contextual detail in the final conclusion, but on the whole this was well answered.

Question 6

There were many good attempts at the final question on this paper and nearly 40% obtained full marks here.

The demand in part (a) required candidates to use algebraic integration and this instruction was missed by a significant minority. These candidates unfortunately scored no marks in part (a). Some evaluated $\mathrm{E}(X^2)$ correctly and then stopped, scoring two marks out of four. Others used the correct variance formula but obtained an incorrect value of $\mathrm{E}(X^2)$ meaning that both accuracy marks were lost.

There were two types of method for part (b). Some candidates evaluated one or more integrals of the form $\int_a^b \mathrm{f}(x)\mathrm{d}x$ for suitable values of *a* and *b*. Many neglected to include the probability from the second line of the probability density function arriving at an incomplete solution. It was not uncommon to see the integrals $\int_{0.5}^{\frac{11}{3}} \frac{1}{4}\mathrm{d}x$ or $\int_{0.5}^{\frac{11}{3}} \frac{1}{8}(x^2+2x+1)\mathrm{d}x$ neither of which plays any part in a correct strategy since two separate integrals with different integrands are needed to determine the required probability using this method.

The second approach consisted of finding the cumulative density function F(x) first and then evaluating F(x) for suitable values of x. For those opting for this method, full marks could only be gained from correct cumulative distribution functions. Many candidates arrived at cumulative distribution functions which lead to negative probabilities and should have realised that an error had been made.

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