Examiners' Report
Principal Examiner Feedback

Summer 2019

Pearson Edexcel International A Level

In Statistics S2 (WST02/01)

Edexcel and BTEC Qualifications

Summer 2019
Publications Code WST02_01_1906_ER

Statistics S2 (WST02)

General introduction

The paper was accessible to all candidates and overall there were some strong performances. Candidates sometimes struggle when translating certain topics in context into correct probability statements. There were a few questions where candidates used calculators and lost marks as a result. A request to use algebraic integration means that full working must be shown. The same goes for "show that" questions where every step of the working must be shown.

Comments on Individual Questions

Question 1

Part (a) was well answered with the majority of candidates showing their method clearly. A minority wrote down a few numbers and then the answer without showing how the answer came for the stated numbers. In part (b) there were plenty of correct non-contextual responses which gained B0 B1 but not many where 2 assumptions were given with the required context. Part (c) was well answered. The main errors were in (i) not to give the answer to an appropriate degree of accuracy (3 significant figures) and in (ii) to rewrite the inequality as $P(X \leq 8)$ or $1 - P(X \geq 8)$ rather than $P(X \leq 7)$. In part (d) few candidates wrote down the distribution required. This could be implied by a calculation of the correct form however, many of those who did not write B(3, "0.2687") forgot the ${}^{3}C_{n}$ resulting in the loss of the marks. Part (e) was a straightforward normal approximation to the Poisson distribution. The main error was not using or using an incorrect continuity correction.

Question 2

Generally there was a mixed response to this question. It was either not attempted or gained full marks. Of those responses which gained full marks some seemed to require no working and others displayed excessive amounts of working. In part (a) if the student identified (0, 0) and (1,-1) as the two samples that resulted in $T = 0$ then they were almost always able to explain their reasoning clearly. There were a few who thought (0, 0) was the only way to obtain $T = 0$ and finding P (0, 0) = 0.16, realised it was incorrect and simply multiplied by 2.
Part (b) was well done by those who attempted the question. The main errors were giving the samples with the associated probabilities or incorrectly calculating the sampling distribution of the mean or even the range.

Question 3

For many candidates this was good source of marks with many correct solutions. In part (a) the majority of candidates found the equation from the given $E(X)$ and $Var(X)$ of the continuous uniform distribution but a number of candidates failed to find two correct equations. The most usual mistake at this stage was to write $\frac{1}{2}(b-a)^2 = 3$ or $\frac{1}{12}(b+a)^2 = 3$. Many candidates who did have two correct equations made simple sign errors or failed to use brackets correctly. In terms of the approach used to solve the equations the majority tended to set up two linear equations and solve by eliminating one of the variables, but scripts were seen forming a quadratic. In part (b), those candidates who gained full marks for part (a) had no problems in deducing that the $P(X > 6 + \sqrt{3})$ was required. However, incorrect answers to part (a) really caused problems for candidates attempting part (b). They spent a lot of time wrestling with awkward equations using unpleasant values from (a). A few scripts were seen with the correct final answer still in surd form. Where the answers here were incorrect it was usually because they forgot the bracket around the $(6 + \sqrt{3})$, thus resulting in the final answer of $\frac{3+\sqrt{3}}{3}$. Another common error was to work out $P(X > \sqrt{3})$. In part (c) many candidates were successful in writing out the cumulative distribution function. However, a common error was to mix the letters x and y when writing out the cumulative distribution function or writing $\frac{y}{6}$ instead of $\frac{y-3}{6}$.

Question 4

The majority of candidates found part (a) of this question straightforward. The most successful method was calculating the probability and comparing it with 2.5%. Those who attempted to use 97.5% were less successful and this is not a recommended route for these tests. Most candidates knew how to specify the hypotheses with most candidates using 2.5 rather than 10. Some candidates used p, or did not use a letter at all, in stating their hypotheses, but most of the time they used λ. A minority found $P(X = 6)$. If using the critical region method, not all candidates showed their working clearly. A sizeable minority of candidates failed to put their conclusion back into the given context. Reject H_0 is not sufficient. Only a minority of responses seen had conflicting statements such as reject H_0 matched with no change in rate of accidents. In part (b), candidates either gained full marks or no marks. There were several scripts where candidates had failed to secure any marks in part (a), but then correctly performed the hypothesis test in part (b). The majority of candidates knew that a 1-tailed test was needed in (b) but the question asked how the test would differ from the one in (a). If a student had used a 1-tailed test in (a) in part (b) they needed to identify that the test would not change [apart from the conclusion]. This rarely happened resulting in the loss of the B mark. They could however gain the M1A1

Question 5

Part (a) proved to be a challenge for most candidates. The most common incorrect method was differentiating to find the maximum (not recognising that the extremes of the x values could be the maximum value) or setting the function equal to zero. Those who recognised the limits as the maximum values were generally successful with this part.

Most candidates were successful with part b (i) although some were not able to cope with the 'k' and so the curve touched or crossed the x-axis. In part b (ii) the 'k' once again caused a few problems with the most common incorrect answer being (2, 1). Part c was a show that question and whilst most candidates could integrate correctly and reach the correct solution, many dropped marks by not demonstrating the substitution sufficiently for a 'show that' question. Most candidates were successful in part (d) gaining full marks. However, poor notation resulted in some candidates losing 4 marks through writing $x\left(x-2\right)^2+1$ and then forgetting to multiply the 1 by x. As in part c, the most common error in part (e) was in not demonstrating the substitution clearly enough to warrant the marks. M1A0 was a common mark here - a correct algebraic statement but then no numerical substitution.

In part (f) many candidates were able to find a suitable argument for this, often based around $\mathrm{P}\left(X < 1\right) = \frac{5}{9}$. candidates who formed an equation and calculated the median were also usually successful.

Question 6

In part (a) many candidates had clearly learned set phrases for the requirements for a binomial model and didn't think about adding context to their comments. Comments about pots cracking and pots firing were mixed and a lot of candidates commented that there should be only two outcomes (cracked or not cracked) which was not awarded marks as the question stated that they either cracked or did not crack so was not an assumption. Part (b) was well answered. In part (c) candidates who used the formula for a Binomial value were more successful than those who used tables. Most candidates were successful in gaining the marks in part (d) and showed an understanding of the difference between greater than or less than for discrete random variables. Many candidates opted for the $\mathrm{P}(X \leq k - 1) = 0.9887$ approach. Errors occurred mainly when candidates apply an algebraic form using $k+1$ and k which caused some confusion. In part (b) the majority of candidates stated the hypotheses correctly but lost marks by not stating or using the correct binomial model. A minority of candidates wrote their critical region in the form of a probability. In part (f) the majority of errors occurred when the student had identified a two tailed critical region in part (e) and the correct answer of 0.355 was added to another probability.

Question 7

This question afforded easy marks for those who had a good understanding of the topic and who were able to present their work clearly. Almost all candidates were able to start this question with an integration using the correct limits of 1 and x. A minority used an indefinite integral with $\mathrm{F}(1) = 0$ but both of these methods generally resulted in the correct answer. The third line of the CDF was more challenging. Errors included integrating $\frac{3}{32}x\left(x-4\right)^2$ using the limits 2 and x but then forgetting to add F(2). In some cases the correct method was used but because their first part was wrong they needed to show their substitution clearly to find F(2) and without doing so lost them marks. The use of an indefinite integral with F(4) = 1 was a more straightforward method and proved more successful. The mark for the 1st and the 4th lines of the CDF were usually gained. There was a mixed response to part (b). Sometimes it was very well done with candidates clearly well versed in solving this type of problem. Most who correctly found F(2.165) and F(2.175) were able to explain why this showed that the median = 2.17 to 2 d.p. Others used their calculator to solve F(m) = 0.5 and were also able to explain correctly why m = 2.17471… meant that the median rounded to 2.17. However many who used their calculator did not consider writing down a more accurate solution than 2.17 and so could not gain the credit. However many candidates wrote F(2.17) = 0.49679… ≅ 0.5 and incorrectly thought that this was an adequate solution.

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