Examiners' Report/
Principal Examiner Feedback

January 2014

Pearson Edexcel International A Level in Statistics 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

January 2014
Publications Code IA037875

Statistics S2 (WST02)

General introduction

The paper was accessible to the majority of candidates and most seemed to have adequate time to attempt all question. Questions 1 and 5 allowed all candidates the opportunity to score a number of marks whilst Questions 4 and 7 were good discriminators at the top end. Mathematical techniques (solving quadratics, differentiation and integration) were generally applied with confidence though candidates were sometimes let down by their inability to select the appropriate statistical technique or give appropriate contextual responses.

Report on Individual Questions

Question 1

This question was answered very well by almost all candidates and most displayed a confident start to the paper by picking up full marks here. In part (a)(ii) some candidates used $\mathrm{P}(X \leq 1)$ but this was very much a minority.

In part (b) candidates using a Poisson approximation invariably got the correct answer. A few used greater than 10 rather than at least 10 but fully correct responses were seen often.

Question 2

Though this question was generally well answered overall, correct responses to parts (a) and (b) were rarely seen. Many candidates focused on opinions rather than the customers and few indicated any sort of list for part (a). Some candidates tried to give textbook definitions of ‘sampling frame’ and ‘sampling units’ without making any contextual reference to the customers at Bill’s restaurant.

Part (c) was done well by the majority of candidates and many were able to give a correct advantage and a correct disadvantage of taking a census. Some responses were not appropriately labelled whilst others focused on taking a sample survey rather than a census.

The attempts at the hypothesis test in part (d) were generally done well and many fully correct solutions were given. Most candidates correctly used p for the hypotheses but some used l or indeed gave no letter. Candidates had little difficulty identifying the correct binomial distribution, B(50, 0.3), but a significant number calculated $1 - \mathrm{P}(X \leq 20)$ rather than $1 - \mathrm{P}(X \leq 19)$. Those candidates opting to use the critical region had mixed success. After identifying $\mathrm{P}(X \geq 21) = 0.0478$, many compared this value to 0.05 to incorrectly reject the null hypothesis. A correct contextual conclusion was given by a large number of candidates who have been trained to use the wording given in the question but this mark could only be scored following a fully correct solution. Still a significant minority simply offer no context to their conclusion.

Question 3

Candidates' understanding of the cumulative distribution function alongside correct differentiation and integration allowed many of them to score full marks on this question. However, slips and carelessness sometimes led to the loss of many marks here.
Most candidates were able to form the equation correctly in part (a) although a number equated the expression to 0.4 or other values. If the equation was correct, nearly all candidates were able to solve the quadratic correctly to obtain a result to the required accuracy. A significant minority of candidates simply left this part blank but went on to be successful with the rest of the question.

Those who were able to differentiate F(x) correctly were generally able to score full marks in part (b). A common mistake was to not multiply out the brackets correctly which led to incorrect differentiation of the final term. It was not uncommon to see xF(x) used to find E(X) losing all the marks for this part of the question.

Those who found (b)(i) correctly had good success with finding Var (X) in (b)(ii). Mistakes such as only finding E(X^2) or subtracting m instead of μ^2 were seen only occasionally.

Question 4

This question showed that many candidates are still not confident with critical regions and significance levels. Some good attempts were seen, with part (c) discriminating the most able candidates.

Most candidates were able to state H_1: $\lambda > 1.5$ as the alternative hypothesis in part (a). The most common errors were using another number instead of 1.5, writing $\lambda \neq 1.5$ or missing this part altogether.

Nearly all candidates used the Poisson distribution for part (b) of this question, although a few continued to use Po(1.5) instead multiplying 1.5 by 4 giving $\mu = 6$ and using $C \sim$ Po(6). Others also lost marks for stating $P(C > 10) = 1 - P(C \leq 9)$. A number of candidates achieved the correct answer of 0.0426 (4.26%) but then went on to state "the significance level is 5%" thus losing the final accuracy mark in this part.

Candidates used various methods for this part of the question. They needed to look at the values of $P(X \leq 10)$ for different values of μ (e.g. 7 and 7.5). Many of those who correctly stated that $\mu = 7$ failed to then divide this by 4 to achieve $\lambda = 1.75$, or equivalent.

Question 5

This question allowed candidates to pick up a good source of marks with only part (c) causing many candidates difficulty. Nearly all candidates realised that a Poisson distribution was appropriate for part (a). However, a few candidates then used Po(15) instead of dividing by 12 to get the monthly rate of 1.25 and hence $X \sim \text{Po}(1.25)$. The use of the formula was generally correct with $\text{P}(X = 3) = \dfrac{\text{e}^{-1.25}1.25^3}{3!}$, but many answers were not given to the required accuracy of 3 significant figures.

As the answer was given in part (b), it was essential that all working was shown. It was necessary to state either $\text{P}(X \geq 2) = 1 - (\text{P}(X = 0) + \text{P}(X = 1))$ or $1 - \text{P}(X \leq 1)$ and also show a correct expression such as $1 - \text{e}^{-1.25}(1 + 1.25)$ which should be stated in terms of e prior to using the calculator and converting to a decimal.

Part (c) seemed to trick even the most able candidates. Instead of calculating 0.355^4 as was required, many ignored the keyword ‘each’ in the question incorrectly turning this into a Poisson distribution with candidates attempting Po(5) and calculating $1 - \text{P}(X \leq 2)$.

For part (d) of the question those identifying the Binomial distribution, B(12, 0.355), were able to make good progress here. Some still tried to use $n = 15$ in this part. Errors with inequalities such as using $1 - \text{P}(X \leq 2)$ were also common. Candidates should take care when typing long expressions into their calculators as a number of correct expressions were seen followed by an inaccurate answer.

Question 6

Most candidates were able to access this question. In part (a) the standard of sketching varied and a large number of candidates were unable to distinguish between a quadratic and a linear function. A large majority of candidates were able to recognise $k(6 - 2x)$ as a line segment, but $k(x + 1)^2$ sometimes appeared as a linear function or a quadratic function with incorrect curvature. The x-axis was usually marked with -1, 1 and 3 in the correct places, but $4k$ was frequently omitted on the vertical axis.

Overall, the attempts at proving that $k = \dfrac{3}{20}$ were very pleasing, with few algebraic errors seen. The majority of candidates expanded $(x + 1)^2$ before integrating, with very few integrating directly as $\dfrac{(x+1)^3}{3}$. Weaker solutions equated the area of each integral to $\frac{1}{2}$ (or 1), yielding different values for k.

In general, candidates knew how to find F(x) in part (c). The first and fourth lines of the cumulative distribution function were almost always stated accurately. Carelessness, however, saw some candidates writing the first line as $\text{F}(x) = 0,\ x < 1$ or the fourth line as $\text{F}(x) = 0,\ x > 3$.

Candidates using an indefinite integration approach with "+ c" had success obtaining the third line of the cumulative distribution function by using F(3) = 1, though many responses lost accuracy in the second second line of the cumulative distribution function either by assuming that $c = 0$ or making arithmetical errors plugging in –1 into the equation F(–1) = 0.

Those opting for a definite integration approach were generally successful with the second line of the cumulative distribution function, but forgot to add F(1) to the integration of the second section of the probability density function.

In part (d), a large majority of candidates knew how to approach finding the median, though some were unsure how to proceed and setting both the second and third lines of F(x) equal to 0.5 was not uncommon. Weaker candidates simply evaluated F(0.5). Those who correctly equated the third line of F(x) to 0.5 were virtually always able to solve their quadratic equation correctly.

Question 7

Many candidates found the final question very demanding, but persevered and despite the original nature of the question, many fully correct responses were seen.

The solution of two simultaneous equations to find the mean and standard deviation of a normal distribution should have been familiar from S1. However, applying this concept to a normal approximation of a binomial distribution may have been unfamiliar to many candidates.

The majority of attempts at employing continuity corrections were successful. Obtaining the z-value of 0.75 was usually done well, although a sign error in $z = -1.25$ was often seen.

Attempts at standardisation were generally good, although weaker candidates equated to $1 - z$ instead of z.

Candidates who used μ and σ in correct simultaneous equations generally went on to complete the question successfully. However, those who used np and $\sqrt{np(1-p)}$ frequently became confused in their elimination of one of the unknowns.

A surprisingly large number of candidates, for this level, tried to standardise with the variance, σ^2 or $np(1-p)$.

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 Edinburgh Gate, Harlow, Essex CM20 2JE