Examiners' Report

Summer 2015

Pearson Edexcel International Advanced Level in Statistics S1 (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 2015
Publications Code IA041218

Mathematics Unit Statistics S1

Specification WST01/01

General introduction

On the whole, students found this paper challenging. Although they were generally well prepared for most of the topics on this paper, it was evident by a number of blank responses that some students did not understand how coding should be used with linear regression in question 2. Students must be reminded that probabilities should be given as exact fractions whenever possible and exact values should always be used in calculations to avoid losing accuracy.

Report on Individual Questions

Question 1

This question proved a difficult start to the paper for many with over 40% of students scoring 0 marks. In part (a), by far the most common error was for students to confuse $F(x)$ for $P(X = x)$ and obtain 1/30 as k. Moreover, when this error was made, it was usually repeated in part (b), with probabilities of k, $4k$, $9k$ and $16k$ appearing. Even when the value of k had been found correctly, some students were only able to obtain the correct value for the first two probabilities as many incorrectly calculated $P(X = 3)$ as $F(3) – P(X = 2)$.

Question 2

Though most students were able to make some progress with parts (a) and (b), a number of blank responses seen in (c) indicated that coding is a topic unfamiliar to many. Only the most able students successfully managed the remaining parts of this question. In part (a) nearly all students used the formulas to obtain the correct values of S_{xy} and S_{xx}. Most then went on to make a correct attempt at finding the values of a and b for the linear regression equation. Students must appreciate the requirement in Statistics to give answers to 3 significant figures and that using rounding values in calculations will likely lose accuracy marks, as was certainly the case in part (b) with 0.738 a very common error.

A variety of attempts were made at part (c) and even those who applied a correct method often lost accuracy due to poor rounding from part (b). Some students tried, with very limited success, to start over and calculate S_{sv} and S_{ss}.

Of those that found an equation in part (c), most substituted $s = 130$ into it to at least score a method mark in part (d).

The interpretation of the gradient in part (e) was not well attempted as many simply described the positive correlation.

The most common mistake in part (f) was to substitute $s = 31$ into the regression equation. Of those that did get a correct answer here, some did so via a circuitous route involving the difference between 2 separate calculations.

Question 3

Though most students were successful with part (a) of this probability question, the conditional probabilities required in the remaining parts were more demanding. In (b) most students understood that they only needed to consider the 25 administrators employed less than 4 years rather than all 90. In (c), however, some mistook less than 10 years for between 4 and 10 years and gave a common incorrect answer of $\frac{20}{34}$.

Many students are familiar with the different tests for independence and there were a number of successful attempts at part (d). Too often marks were lost here due to incomplete labelling, many opting to leave the events in terms of A and B without explicitly defining them.

Question 4

This was another question on probability with an accessible start and a challenging end. Some drew a tree diagram to get a better sense of the question and this generally enabled them to complete parts (a) and (b) correctly. A surprising number of students evaluated $\frac{1}{20}\times\frac{1}{20}$ as $\frac{1}{40}$ in part (a). In part (b), not everyone appreciated that there were two possible ways that this could be done.

Most realised that 3 products were required in part (c) though equivalent products were not always given. Those who realised the most efficient approach to part (d) was to find the probability that all 4 of the beads were yellow and take this away from 1 were most successful. When students attempted to find all of the different ways that at least 1 green bead could be selected, they often failed to find all of the 15 favourable outcomes and inevitably scored no marks here.

Question 5

This question was one of the most successfully answered questions on the paper despite the usual difficulties students have with the normal distribution. Over ¼ of the students scored at least 11 marks here. In part (a) most sketches were reasonably drawn with the mean of 55 commonly labelled at the centre of the x-axis. Those not scoring marks here included sketches drawn as semi-circular, not symmetrical or ones which crossed through the x-axis.

The majority of students were able to obtain a correct answer in part (b), although some gave the percentage of the cars that were travelling slower than the speed limit as their final answer (77.34%). In a few cases 75% or 67% was given as the final answer where students incorrectly used the z-value as a probability.

Students found part (c) more demanding but it was pleasing to see that a number of students did use 2.3263 correctly as their z-value (rather than less accurate values between 2.32 and 2.35). Equating the standardisation to a probability continues to be a common mistake for weaker students. Some students attempted a verification method by standardising 102, however, this invariably involved nothing more than standardising 102 alone and hence did not prove the given result. Verification attempts required clear evidence that 102 gave a probability closer to 1% than standardising 101 or 101.5; hence a score of only one mark was usually obtained for these attempts.

Only the most able students made progress with part (d) as errors were often made in the first line of working e.g. writing $P(70 < X < m)$ as $P(X > 70) - P(X < m)$. Of those correctly finding the expression for $P(X < m)$, many again set their standardisation equal to this probability rather than the associated z-value. Incompatible signs between standardisations and z-values were at times seen.

Question 6

Though a number of good attempts were seen in this question, some students did find this difficult and marks were varied. Most were able to answer part (a) correctly but the most common error was to give an answer of ¼ instead of ¾. Part (b) was again generally answered well, but only a minority used the formula that applied here for the discrete uniform distribution, $E(X) = \frac{n+1}{2}$.

As the answer was given in part (c), it was not enough to write down $E(X^2) = 7.5$ without any justification. Again only a small number of students quoted the formula for the variance of a discrete uniform distribution and used it.

Despite showing an understanding of the discrete uniform distribution in parts (a)-(c), many failed to understand the same distribution applied in (d) and wrote down incorrect or inconsistent probabilities. In fact, it was common to see no marks scored in parts (d), (e) and (f).

Able students knew the properties of variance and gave $Var(Y) = k^2Var(X)$ in part (e) although quite a few forgot to square the k when extracting it. Those who went back to basics and used $Var(Y) = E(Y^2) - [E(Y)]^2$ were invariably less successful and many of them forgot to square the $E(Y)$ thus losing both marks.

Part (f) was more often than not left blank and was generally only scored by those who had secured full marks to that point. Some did seem to spend a significant amount of time and work including setting up and solving simultaneous equations to get there though.

Question 7

This was the most successfully answered question on the paper with over ½ of the students scoring at least 5 marks here. The vast majority were able to gain both marks for the plotting of points on the scatter diagram in part (a) with only a few misreading the scale.

Part (b) was answered well though students sometimes needed to be clearer in their written expression. Comments such as 'it is not in range' were considered too vague as it was not sufficiently clear that they were referring to the blood protein of *H*.

Whilst the majority substituted correctly and obtained the correct answer to what should have been a very straightforward 2 marks in part (c), a significant number of students seemed to believe $423\frac{5}{7}$ meant $423 \times \frac{5}{7}$ and gave a common incorrect answer of 0.959. Another issue here was to give a truncated answer, 0.809, rather than a correctly rounded one.

Many students incorrectly relied on the stock answer '*r* is not affected by coding' in part (d). Those indicating that it would increase, presumably did so thinking that including *H* meant the regression line would get steeper.

Question 8

This question produced a good spread of marks with only the most able students persevering through all parts here. Most realise the requirement for the data on a histogram to be continuous and scored in part (a).

In part (b), many either obtained a relationship between squares and people or labelled the frequency density on the *y*-axis of the histogram. A common mistake was to assume that the frequency density was constant in the range 18 to 25 minutes rather than splitting it into two separate areas.

The most common mistakes in part (c) included using 25.5 as the class boundary or misreading the height of the bar and obtaining a frequency of 35 in the group. This mistake was repeated in (d) where often the 3rd and 5th frequencies were incorrect in the calculation. This should have been a sign to students that a mistake was made as the answer 6390 was given.

A significant number of students did not answer parts (e)-(g) indicating they either ran out of time or failed to read the questions on the final page. Of those attempting part (e), most did so correctly. The mark in part (f) was generally only scored by those who achieved full marks on the entire question as this value was dependent on the previous parts being correct.

For those who persevered to the end, many did appreciate that their answer in part (f) determined whether or not the normal distribution was appropriate to model these data.

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, United Kingdom