Examiners' Report

Principal Examiner Feedback

October 2020

Pearson Edexcel International GCE

In Statistics 1 (WST01)

Paper : 01 Statistics S1

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

October 2020
Publications Code: WST01_01_2020_ER

General

This paper was accessible to all candidates with all of the questions having opening parts that all prepared candidates should have been able to engage with.

Report on individual questions

Question 1

This proved to be a good introductory question to the paper with many students scoring full marks and few failing to score anything at all. Some did not realise they could find the value for k and simply found $\mathrm{E}(X) = \frac{41}{k}$ and a few, who did use the fact that the sum of the probabilities = 1, made a slip in solving $\frac{25}{k} = 1$ and gave their value of k as $\frac{1}{25}$

Question 2

The tree diagram was usually completed accurately in part (a) and most were able to use their diagram to find an expression in p and q for P(Water) in part (b), there were some careless errors made in simplifying this expression though such as 0.25×0.1 being given as 0.25 though fortunately this did not affect the mark here. Part (c) proved more challenging with most attempting $\mathrm{P}\left(V \cap W\right) = \mathrm{P}(V) \times \mathrm{P}(W)$ using their answer for $\mathrm{P}(W)$ from part (b). There were numerous errors occurring here with equations such as $0.4 = 0.25 \times \left(0.4p + 0.35q + 0.1\right)$ appearing quite often. Those who did apply this rule successfully often made careless arithmetic errors in simplifying their equation which had implications for part (d). It was disappointing that very few students realised that if W and V are independent then they could simply write down $\mathrm{P}(W) = 0.4$ from the tree diagram.

Part (d) revealed that many students' grasp of conditional probabilities was shaky. Those who did understand this topic were often able to secure the first two marks for forming an equation in p and q using the given probability. They were then left with the daunting task of solving two, unpleasant, linear equations in p and q. Of course the question was much simpler: those who realised that $\mathrm{P}(W) = 0.4$ and therefore $\mathrm{P}(J) = 0.6$ could use the given probability to quickly obtain $1 - q = \frac{2}{3}\mathrm{P}(J) = 0.4$ and find q and then p very easily. Few fully correct solutions for p and q were found. In the final part few used conditional probabilities as intended though they were sometimes able to make a correct conclusion based on their values.

Question 3

The standard normal distribution calculations in parts (a) and (b) were answered quite well but parts (c) and (d) was more challenging.

In part (a) the vast majority could standardise correctly and obtain a probability from the tables. However a small number were still not sure what to do with this value and failed to subtract it from 1 and some poor rounding (0.555… rounded to 0.55 rather than 0.56) meant that others had a slightly inaccurate value and lost the final mark. Many made good progress with part (b) though many of the usual errors were still seen. Some standardised and set equal to 0.8 (from the probability of 80%) rather than a suitable z value. Others found the z value of 0.8416 from tables but then used 1 – 0.8416 in their equation. Those who did use a z value of 0.84 or better could score at least one mark here but many did not realise that a negative value of z was required and an answer of 4.56 was common. A few attempted to use their calculators with mixed success; if a calculator is used then they must give evidence of using values at least as accurate as the tables and so an answer of just 3.04 without showing their calculator value of 3.0425…first was insufficient to secure all 3 marks. In part (c) few students realised that a conditional probability was involved but a number did arrive at a probability statement of the form $\mathrm{P}(D > g) = \frac{1}{3}$(a) and some of these went on to find a correct value of g. A significant minority confused medals with certificates and so we had $\mathrm{P}(D > g) = \frac{1}{3} \times 0.80 = \frac{4}{15}$ but a special case was available on the mark scheme and those who arrived at $g = 4.36$ were still able to obtain 2 marks here. Part (d) of course did not require a knowledge of the normal distribution but simply some careful thought about the arrangements for the group of 3 medal winners. Some did give a probability of the form $1 - p^3$ and a few obtained the correct answer.

Question 4

Part (a) was answered very well with nearly everyone finding the correct value for $Q3$ and most finding the outliers correctly. A few thought that the lower outlier limit was $1.5 \times \mathrm{IQR} - Q_1$ and some thought the upper limit was $1.5 \times Q_3$ In part (b) most drew the box plot correctly with a majority having whiskers going to 10 and 45 though some used 9 and 49. Part (c) was answered very well indeed and most scored both of the marks here though a small minority gave a correct comparison but then concluded that the data was positively skewed. Most students failed to recognise the change in IQR and therefore the need to re-calculate the limits for the outliers. Those who did show this step invariably completed the box plot correctly. The final part was rarely answered correctly and required some careful reasoning. For the median to be the same, the values must be either side of the median, for the lower quartile to increase then the lower value must be inside the left hand said e of the box. For the upper quartile to remain the same the values must be either side of the upper quartile and so the other value must be in the upper whisker.

Question 5

In part (a) most realised that they simply needed to substitute $x = 80$ into the given equation. However some then failed to give an answer for the rent in dollars and they either missed including a \$ symbol or failed to multiply their answer of 16.946 by 1000. Part (b) was probably the easiest part of the paper and nearly all students scored full marks here. Most students knew how to find the correlation coefficient r and realised that they required a value for S_{xy} but far fewer were able to find this successfully. Those who realised that it was simply the gradient of the regression line multiplied by S_{xx} usually went on to score all 4 marks here. A few students gave their final answer as 0.91 rather than to 3 significant figures as instructed on the front of the question paper. Part (d) proved, unsurprisingly, challenging. A good number of students were able to show that the new mean and the old mean were both equal to 90 but few went on to state that since the new value of $x = 90$ then the extra $\left(x-\bar{x}\right)$ term would be zero and so S_{xy} was unchanged. Some students recognised in part (e) that the gradient of the regression line would be the same and those who also calculated the new value of $\bar{y}$ were able to obtain a correct new equation. In part (f) most stated that 3000 m^2 was too large an area to use but few gave a suitable reason by either comparing 3000 m^2 with the mean area of 900 m^2 or comparing 300 with the mean of $x = 90$.

Question 6

Parts (a) to (d) here were familiar territory for many students and there were many good answers to these parts. The usual errors in (b) were very occasionally seen: thinking Var(A) = E(A^2), subtracting E(A) instead of [E(A)]^2 and rarely dividing a correct expression for E(A^2) by 5. There was still some confusion over the correct terminology in part (c) with "discrete random variable" being quite common whilst other invented their own title such as "fair distribution" or "even distribution". Most students obtained $k = 6$ in part (d) by calculating E(B), it was extremely rare to see a student using "symmetry" as their reason.

The final 3 parts of this question were, as expected, more challenging. In part (e) many attempts found the z value for Sam as $-\frac{1}{6}$ but few realised that for Tim this meant that $\frac{3.5-A}{4}<-\frac{1}{6}$ and therefore $A > 4.2$ leading to a probability of $0.25 + 0.15 = 0.4$ In part (f) most students gave their value of a and their value of b either directly or in a suitable standardisation and then we could easily see whether they were working with a correct z value of -3.5 It was unfortunate that those who simply looked up 3.5 in the tables would see a value of 0.9998 and without supporting working this could not be awarded full marks. Those who clearly had used $A = 7$ and $B = 1$ were often able to score the final 2 marks in part (g) but some failed to relate this part back to the earlier distributions and thought some complex calculation involving a normal distribution was required here.

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