Pearson

Examiners' Report

Principal Examiner Feedback

January 2017

Pearson Edexcel International A-Level

Statistics S1 (WST01)

edexcel

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 2017
Publications Code WST01_01_1701_ER

Introduction

There seemed to be a number of candidates taking this paper who were unaware of our usual conventions that probabilities should be given as fractions or decimals rather than percentages. On this occasion we did allow percentages in several places but in general probabilities should be given as a number in the range [0, 1]. The drawing of a tree diagram in question 5 proved troublesome for a number of candidates and many seemed unfamiliar with the usual system of labelling (events at the end of branches and probabilities along the branches). A number of candidates did not seem well prepared for question 6 involving the normal distribution.

Comments on individual questions

Question 1

Even some candidates who found the correct answer to part (a), were clearly not confident about the important relationship between area and frequency for a histogram and only achieved their answer after half a page of working. The most efficient solutions were from those who correctly labelled the frequency density axis on the histogram or those who identified that 1 large square represented 5 tomatoes. Many errors were seen though, throughout this question, due to carelessness when reading the scales. Part (b) was answered well with many finding the correct frequency for those tomatoes weighing more than 3 grams but some failed to give their answer as a probability. A similar problem arose in part (c) with many giving a frequency of 55 rather than a proportion of 0.55. The most successful approaches used the area to the right of 6.25 and then simply found $0.75 \times 16 + 25 + 10 + 8$. Part (d) proved quite challenging and only the better candidates were able to give a clear argument based on their answer to part (c). Many seemed to "forget" that the 0.55 represented the proportion **above** 6.25 and simply argued that because the median was the 50% point it must therefore be below 6.25. In part (e) we allowed candidates to effectively "start again" and a number found an estimate of the median using linear interpolation from which they were able to establish that the skewness was negative because the median was greater than the mean. In part (f) very few candidates appreciated that selecting 2 tomatoes at random meant that they had to be selected without replacement. A number found the probability of the first tomato being within 0.75 grams of the mean as 0.24 but some answers just stopped here and another common incorrect approach simply squared 0.24 whilst a small minority attempted to use a normal distribution.

Question 2

Part (a) was sometimes left blank or a standard definition of $A \cap B$ was given with no attempt to use the given contexts and a few thought that they had to give $\mathrm{P}(A \cap B)$. Part (b) was usually correct but in part (c) some gave an answer of 12 rather than the probability of $\frac{12}{50}$. The conditional probability required in part (d) was answered well with many simply writing down the correct probability from their diagram and others, who gave a correct ratio expression, picking up at least a method mark if they found one of the correct probabilities. Showing some working is to be encouraged as a numerator of $\frac{5}{50}$ was quite a common error. Only the better candidates used their answers to parts (b) and (d) in part (e) and most chose to prove non-independence by showing that $\mathrm{P}(A) \times \mathrm{P}(C) \neq \mathrm{P}(A \cap C)$. Whilst this did not strictly constitute a correct response to the question, on this occasion, a special case was allowed for 1 of the marks. It appears that many candidates were not aware of the P($A|C$) = P(A) condition for independence. In part (f) some could not identify the correct conditional probability with P(B | C) or P(B | $A \cup C$) sometimes being attempted but many were simply able to write down the correct answer straight from the Venn diagram.

Question 3

In part (a) many candidates found the mean and variance of x rather than y. Those who were using the correct variable sometimes subtracted – 2.25 rather than $(-2.25)^2$. Part (b) (i) was usually answered correctly although some are still losing an accuracy mark because they do not give their final answer to 3 significant figures or better. Whilst there were many fully correct responses to (b) (ii), some candidates failed to appreciate the significance of the negative correlation and thought that, since $|r|$ was close to 1, the value supported Priya's belief. A good proportion of the candidates started part (c) correctly but the usual problems of premature rounding, coupled with errors over the minus signs, meant that only a minority arrived at a final equation with values of both a and b correct to at least 3 significant figures. Minor inaccuracies in part (c) did not affect their answer to part (d) and the vast majority knew what to do here. Many started part (e) correctly too and an encouraging number arrived at a correct equation in w and x but a fairly common error was to write $1.8y + 32 = -0.827 - 0.0339x$ (or their equivalent equation) and they of course made no progress. In part (f) (i) we saw a good number of candidates who multiplied their variance from part (a) by 1.8^2 but some simply multiplied by 1.8 and added 32 and others divided by 1.8^2. Responses to part (ii) were better, with more candidates realising that the correlation coefficient would be the same as in part (b) (i) although some failed to give the numerical value and lost the mark.

Question 4

The first 4 parts of this question were answered very well and these standard calculations involving discrete probability distributions are well known by the vast majority. There were the usual problems in part (b) with some forgetting to subtract 6.9^2 and a very small minority dividing by 4 at some stage. In part (c) some candidates tried $3^2 \times \text{Var}(X)$ but well over 60% of the candidates scored 8 or more marks on this question. Part (e) proved off-putting for some candidates who simply gave up on the question at this point whilst others wrote down the distribution for Y but could not make the connection with X to find P($X = Y$). There were some correct answers seen though sometimes the solutions seemed quite laboured. Part (f) defeated all but the most able and there were few correct solutions seen. Those who started by trying to list the possible cases stood a good chance of success but many seemed unprepared for this type of question.

Question 5

At first sight this seemed to be a straightforward question but over a quarter of the candidates scored 0 and many lost marks in part (a) for failing to label their branches correctly. Since the usual conventions for labelling were clearly not familiar to all the candidates some flexibility was allowed this time but many still lost marks for mixing up the ages and the probability p or simply missing a vital probability or event label. Answers to part (b) were better, with many gaining full marks here even without a correct tree diagram. Some forgot the brackets on their $(1 - p)$ term and this led to an incorrect value for p. In part (c) many did not identify the conditional probability and simply found the probability of being under 50 and using a computer every day but there were plenty of fully correct solutions and indeed over a quarter gained full marks on this question.

Question 6

The normal distribution is still not understood very well by a large number of candidates and over a third failed to score any marks on this question. Part (a) was often answered correctly but many seemed to spend half a page calculating *z*-values before they arrived at their answer. Incorrect answers of 99% or 1 % were not uncommon. Part (b) was a fairly standard "reverse" calculation and many candidates knew how to go about this. Of course some failed to use the percentage points table and had a *z* value of 2.32, 2.33 or 2.34 rather than the 2.3263 or better that was required for full marks but the method and final answer were often correct. Part (c) proved to be beyond most candidates. The first hurdle, where most fell, was to realise that they needed twice the probability of "being in the tail" and the second problem was that few candidates realised the need to square this value. A common error at this stage was to multiply their "tail" probability by 2 rather than squaring it so very few fully correct solutions to this part were seen.

Question 7

Most candidates could make some progress here and the given answer in part (a) provided a helpful opening to the question with most candidates showing clearly the need to add the 4 probabilities together and set equal to 1 and showing an intermediate step before they arrived at the printed answer. Part (b) was a little more challenging and some confused F(3) with $P(X = 3)$ but there were a good number of fully correct solutions here. The final part proved more challenging. Finding the *y* values was often the only mark scored but some thought that the probabilities should be squared. Others tried to use the probability function for *X* with values of x^2 and many ended up with a cumulative probability greater than 1. The very best candidates had little problem in giving a correct table that was clearly labelled but this required the insight that $F_Y(9) = F_X(3)$ etc but few realised this.

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