edexcel

Examiners' Report

Summer 2014

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 2014
Publications Code IA040140

Mathematics Unit Statistics 1

Specification WST01/01

General Introduction

The paper proved to be accessible to all students with Q1, Q2, Q3 and Q5 being answered very well. Q4, Q6 and Q7 proved to be more demanding.

Report on Individual Questions

Question 1

The numerical parts of the question were generally well answered but premature rounding and answers with an accuracy of less than the required 3 significant figures lost marks in Q01(a) and Q01(e). Where comments were required students were less successful. In Q01(c) many did not interpret the correlation coefficient in the context of the question. In Q01(d) some thought a negative value for the correlation coefficient meant the value did not support the researcher's belief. Students should note that fractions were not accepted for coefficients of the regression line in Q01(e). Whilst most answered Q01f(i) correctly by substituting 40 into their regression equation, a large majority lost the mark in Q01f(ii) by failing to realise that this was interpolation or saying 'it' was in the range without specifying age.

Question 2

Students are generally confident at calculating summary statistics such as the median, mean and standard deviation as required in this question. In Q03(a) however, students often struggled to find the correct scale factor for the width and height of the bar. Responses were very mixed, with a substantial number of students working on the assumption that frequency is equal to area rather than frequency being proportional to area.

Q03(b) was largely answered very well. Incorrect answers generally came from an incorrect class width or lower class boundary. There were very few incorrect answers for the mean in Q03(c). The formula for the standard deviation was usually well known. Several students lost time by recalculating $\sum fx^2$ rather than using the given value.

Q03(d) was generally well answered with most students familiar with comparing the mean and median to determine skewness. Students should attempt to use information already obtained to describe the skewness whenever possible as some took the inefficient route of calculating quartiles in order to establish that $Q_3 - Q_2 > Q_2 - Q_1$ when using the mean and median would have been sufficient.

Students had some success with Q02(e)(i) and Q02(e)(ii) explaining the effect of the additional piece of data on the mean and the median but correct responses to Q02(e)(iii) were rare. Many students confused the addition of data with coding and stated that standard deviation is not affected by addition. Several students responded that the values would change but did not actually state the direction of the change.

Question 3

The cumulative distribution function has become more accessible to students and this question was mostly well done with the majority gaining at least 4 marks. In Q03(a) students need to be sure that they are associating the probabilities with their respective *x* values. A significant minority of students attempted to find $\mathrm{E}(X)$ here. In Q03(b) the most common mistake was to use $\mathrm{F}(x)$ as the probability distribution but this mistake should have been picked up as it led to an answer (probability) greater than 1. Q03(c) was the least successful part of this question as many seemed to think that $\mathrm{F}(4)$ meant $\mathrm{P}(X = 4)$ and wrote an answer of 0.

Question 4

This question got a mixed response from students with the most able finding it easy to score the full marks. In Q04(a) a common incorrect response came from confusing probabilities with *z*-values and writing $\mathrm{P}(Y > 17) = -0.6$ as the answer. In Q04(b) the notation caused confusion for some as a number of students failed to appreciate that $\mathrm{P}(Y < \mu) = 0.5$, i.e. that 50% of the area is below the mean no matter what the value of μ is. By this stage of the question many had apparently given up and the conditional probability required in Q04(c) was beyond most students. Those who did write a correct formula $\mathrm{P}(Y < \mu \mid Y < 17) = \dfrac{\mathrm{P}(Y < \mu \cap Y < 17)}{\mathrm{P}(Y < 17)}$ often interpreted this as $\dfrac{\mathrm{P}(17 < Y < \mu)}{\mathrm{P}(Y < 17)}$ and hence lost both marks in Q04(c). Of those who got the correct ratio many reduced their answer to 0.83 (2sf) and lost the accuracy mark.

Question 5

Many fully correct solutions were seen here with only Q05(f) discriminating the most able students. Q05(a) was correctly answered by the vast majority of students and most went on to find the required probabilities in Q05(b). Very few students failed to score both marks for part Q05(c), although the subtraction of the mean rather than the square of the mean lost the marks here for some.

The slightly different presentation of $\mathrm{E}(aX + b)$ and $\mathrm{Var}(aX + b)$ in parts Q05(d) and Q05(e) did not cause too much difficulty for most students, though some resorted to writing out the new probability distribution for *Y* and often made an error in one of the calculations in doing so.

Q05(f) was found to be challenging, with some students not even attempting this part. For those attempting to solve using inequalities, a large proportion failed to simplify their inequality correctly; a commonly seen mistake was to neglect to change the direction of the inequality sign when dividing by –3. Students attempting to change variable and use the distribution for *Y* were generally more successful.

Question 6

This Venn diagram question challenged many students with Q06(d) serving as one of the most discriminating parts of the entire paper. The overlapping regions in the Venn diagram were not well understood by a large amount of students in Q06(a). Whilst most students could specify $\mathrm{P}(D)$, writing $\mathrm{P}(S)$ proved difficult for many and using $\mathrm{P}(S \cap D) = 0.04$ was a common mistake. Students using the multiplication rule for independent events proved more successful than those using conditional probability. Even those who were unsuccessful in Q06(a) managed to pick up both marks in Q06(b) using the fact that sum of probabilities = 1.

Students had more success with the conditional probability in Q06(c)(i) than in Q06(c)(ii). In Q06(c)(i), those who did set up the correct ratio of probabilities sometimes lost accuracy by giving an answer to only 2sf. In Q06(c)(ii), students struggled to identify the correct denominator with many using 0.15 instead of the 0.54 as required.

Most students did not see the link between Q06(c) and part Q06(d) and fully correct solutions in Q06(d) were rare. Many tried to oversimplify this part by attempting $\mathrm{P}(D) \times 63$ whilst others ignored the conditional probabilities and gave $\mathrm{P}(D \cap M \cap S) \times 27 + \mathrm{P}(D \cap M \cap S') \times 36$ as an incorrect attempt.

Question 7

There were many students who succeeded with Q07(a). This was a ‘show that’ question which required a fully correct solution and a large proportion lost the final accuracy mark due to errors in mathematical communication, writing mathematical statements that were false. Students must realise that a ‘show that’ question always requires full working to be shown and that any incorrect statements or poorly explained working will be penalised.

Q07(b) was found to be challening, with only a few getting 0.2578 and many simply using a z-value associated with 0.191. Some students standardised with 32 and 12 seen but then equated to a probability rather than a z-value.

It was rare to see a fully correct solution to Q07(c). Those who recognised the need to multiply the probabilities seldom went on to multiply this by 3. Some drew a tree diagram which helped enable them to understand that there were multiple combinations to consider. A lot of incorrect responses involved adding 0.0668 and (2 × 0.191).

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