Examiners' Report
Principal Examiner Feedback

January 2018

Pearson Edexcel Level 3 Award
In Statistical Methods (AST30)
Paper 1

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

Edexcel Award in Statistical Methods (AST30)
Principal Examiner Feedback – Level 3

Introduction

There was no evidence to suggest that students had difficulty completing the paper in the given time.

The vast majority of candidates completed their answers in the spaces provided and many showed the steps in their working.

Some candidates did not learn all the required formulae for the examination.

It was pleasing to see so many students showing the intermediate stages in their calculations.

Reports on Individual Questions

Question 1

Part (a) was answered well by the majority of the candidates who were able to calculate a probability based upon a completed sample space.

Part (b) was answered well by the majority of the candidates who could correctly use a theoretical probability to calculate an estimate of the frequency of an outcome occurring.

Question 2

Part (a) was generally answered well. Many candidates were able to explain what is meant by categorical data and by quantitative data.

Part (b) was answered well. Most candidates could give an advantage of collecting primary data with references to accuracy, reliability, the data being up to date and collecting the required data all being popular responses. Candidates were also generally able to identify an advantage of collecting data from the internet, although some incorrectly gave a disadvantage of collecting data from the internet and others referred to collecting a range of data which was not an acceptable response.

Part (c)(i) required candidates to give an advantage of taking a sample rather than using a census. The majority of candidates were able to give a correct advantage with reference to collecting a sample being easier being a popular response.

Part (c)(ii) was answered well by the most able candidates, however many responses omitted reference to 'all' when describing the population for the sample and others referred to data sources, for example car dealerships.

Question 3

Part (a) was answered really well by the candidates. The majority of the candidates gave a correct answer.

Part (b) was generally answered well by candidates. The majority of the candidates were able to give an assumption for the capture-recapture technique. Some candidates incorrectly referred to a potential issue, for example 'the paint has rubbed off' and others restated the estimate calculated in part (a).

Question 4

Part (a) was answered really well by the majority of the candidates with most opting to use fractions to complete their tree diagram. The only error observed was to complete the tree diagram using the probabilities for without replacement.

Part (b) was answered really well by candidates. The majority of candidates were able to calculate the required probability and it was pleasing to note that a high proportion of candidates showed their working. Candidates who had worked without replacement in part (a) were generally able to correctly calculate the probability based upon their tree diagram.

Question 5

This question was answered very well by candidates. It was pleasing to see that candidates were familiar with the calculation for the strata size where the population was stratified by two categories.

Question 6

Candidates generally recognised that this question required them to work with the lower quartile and upper quartile to calculate limits for outliers. Around half of the candidates were able to successfully calculate the limits for outliers and demonstrate that there were no outliers in the data as required. A minority of candidates found the limits for outliers, but did not demonstrate that there were no outliers in the data by comparing the lowest and highest values to the outlier limits. Some candidates made errors in finding the quartiles from the list of data and others gave an incorrect formula to calculate the limits for outliers.

Question 7

Part (a) was answered very well. The majority of candidates were able to calculate the chain base index number required.

Part (b) was answered well by many candidates who were able to calculate the geometric mean of the index numbers. Some candidates calculated the *arithmetic* mean of the index numbers rather than the geometric mean.

Part (c) required candidates to interpret the geometric mean in the context of the question. About half of the candidates were able to give a full interpretation. Where full marks were not awarded this was generally due to omitting per year/annual from the answer.

Question 8

Part (a) was reasonably well answered by the majority of candidates. The majority of candidates were able to draw a correct histogram, however some omitted the label on the vertical axis and so gained two rather than three marks. A minority of candidates made an error in calculation of one or more of the frequency densities. Some candidates used an incorrect formula in their attempt to calculate frequency density. It was pleasing to see that the majority of candidates knew how to calculate an estimate of the mean from grouped data. A minority of candidates made arithmetic errors in their calculation, but these candidates had generally shown their calculation steps and therefore were able to gain partial credit.

Part (c) was answered well. The majority of candidates were able to correctly calculate an estimate for the standard deviation of the ages. A minority of candidates used an incorrect formula in their attempt to calculate an estimate for the standard deviation.

Question 9

Part (a) was generally well answered. The majority of candidates were able to correctly complete the back-to-back stem and leaf diagram for the times taken and included an appropriate key. One error which arose was listing the leaves for boys with the values increasing from left to right as would be expected in a single stem and leaf diagram.

In part (b) it was pleasing to see many candidates using the correct language to compare the distribution of the times taken. Many candidates gained both marks from this part. A common error was for candidates to state the median and interquartile ranges without comparison. A minority of candidates chose to compare the highest and lowest values which did not gain credit.

Question 10

Part (a) was not done well. Although the majority of candidates gained at least one mark for an attempt at a box plot there were many errors in the box plots drawn. A common error was to read the cumulative frequency curve at 30, 60 and 90 rather than 25, 50 and 70 to find the lower quartile, median and upper quartile.

Part (b) was answered very well by a minority of candidates. The majority of candidates demonstrated awareness of the required comparisons, but often had errors in the comparative statements made. A minority of candidates stated values without making comparisons, in this case they gained 1 mark if they stated the correct skew for the two distributions.

Question 11

Part (a) was done well by some candidates who accurately calculated Spearman's coefficient of rank correlation for the data. Some candidates lost marks as they made arithmetic errors when calculating the values of d or d^2. It was disappointing to see that a significant minority of candidates made errors in the formula and in some cases gave an answer which was greater than 1 without recognising that this was not possible.

Part (b) was not done well. The majority of candidates stated that there was positive correlation rather than interpreting the answer to part (a) in the context of the question as was required.

Question 12

Part (a) was done well. Most of the candidates were able to describe the trend shown on the time-series graph. Candidates should be advised that the correct word to describe the trend here is downwards, and that e.g. "decreasing" and "negative" are merely condoned.
Part (b) was done well. Many candidates were able to calculate the required mean seasonal variation for quarter 1. A common error was not being able read the correct scale or to give the answer -4 or -4000 rather than 4 or 4000.

Question 13

This question was generally answered well. The majority of candidates were able to correctly calculate the required standardised scores. A common error was to use the mean and standard deviation for the female data when calculating the standardised score for Greg.

Question 14

Part (a) was answered well. Most candidates were able to complete the Venn diagram correctly.

Part (b) was answered well. In part (i) most candidates were able to correctly find the probability that the student studied all 3 types of dance. Part (ii) caused more difficulties with a common error being to omit the 7 candidates who studied all three types of dance and give the answer $\frac{6}{130}$.

Part (c) was answered quite well. Many candidates were able to calculate the required conditional probability. Where the correct answer was not seen there were a range of different errors observed.

Question 15

The majority of candidates were awarded 1 or 2 marks in part (a). A minority of candidates simply found the z value and did not look this up in the tables. A common error was to find 1 – 0.9918 and give the final answer as 0.0082

Part (b) was not answered well. Common errors were 0.9918-(1-0.8413) giving an answer of 0.8331 and 0.9918+(1-0.8413) which gave an answer greater than 1 which did not appear to worry the candidates concerned.

Question 16

This question was answered poorly. The majority of candidates gained only 1 mark in part (a) as they worked with the mean values rather than finding $\sum x$ and $\sum y$. Another error was to omit the square root in the calculation of r.

Part (b) was not well answered. Many candidates had a value greater than 1 from part (a) but interpreted this as indicating positive correlation in (b).

Question 17

Part (a) was not well answered. The most common error was to add the probabilities of *A* and *B*, but then not to subtract the probability of *A* or *B*. Other errors included multiplying the probabilities of *A* and *B* which was sometimes followed by subtraction of the probability of *A* or *B* and sometimes the product was subtracted from the probability of *A* or *B*.

Part (b) was also not well answered. There were a range of incorrect calculations which utilised the probabilities given in the question and the answer obtained in part (a). It was disappointing to see candidates giving probability answers which were greater than 1 or negative seemingly without identifying that these were not possible.

Part (c) was generally answered well. The majority of candidates were able to correctly identify that the events as mutually exclusive and independent based upon the information given. Incorrect answers included reversing the answers to parts (i) and (ii). Other incorrect answers were exhaustive and intersect.

Question 18

Part (a) was generally answered well. The majority of candidates were able to sketch an appropriate normal distribution curve for the speeds on road *Y*. Common errors included drawing the curve with the same height as that for road *X* and using incorrect start and end points for the curve.

Part (b) was generally answered well. The majority of candidates were able to correctly compare the two distributions.

Question 19
Many candidates answered part (a) well, many candidates could recall the formula for the binomial distribution and apply it correctly. A common error was to calculate $^{10}C_2 10^2 (1 - 10)^{10-2}$.

Part (b) was not answered well as many candidates just worked with *X*=3 and calculated 1-P(*X*=3).

Summary
Based on their performance on this paper, students are offered the following advice:
- Read the question fully and carefully before attempting to answer it.
- Show working out to support the final answer.
- Know how to calculate the standard deviation.
- Be both precise and explicit in comparisons of distributions.
- Check to see if answers make sense in the context of the problem.
- Check to see if their answers are reasonable in the case of probabilities and correlation coefficients.