GCE

Statistics (MEI)

Advanced Subsidiary GCE **AS H132**

Report on the Units

June 2010

H132/R/10

This report on the Examination provides information on the performance of candidates which it is hoped will be useful to teachers in their preparation of candidates for future examinations. It is intended to be constructive and informative and to promote better understanding of the specification content, of the operation of the scheme of assessment and of the application of assessment criteria.

Reports should be read in conjunction with the published question papers and mark schemes for the Examination.

OCR will not enter into any discussion or correspondence in connection with this report.

Any enquiries about publications should be addressed to:

OCR Publications
PO Box 5050
Annesley
NOTTINGHAM
NG15 0DL

Telephone: 0870 770 6622
Facsimile: 01223 552610
E-mail: publications@ocr.org.uk

CONTENTS

Advanced Subsidiary GCE Statistics (H132)

REPORTS ON THE UNITS

Chief Examiner's Report

The Principal Examiners' reports that follow discuss the candidates' performances on the individual modules. There is one matter that should be discussed in a general way as it applies to all the statistics modules. This is in respect of arithmetical accuracy in intermediate working and in quotation of final answers.

Most candidates are sensible in their arithmetical work, but there is some unease as to exactly what level of accuracy the examiners are expecting. There is no general answer to this! The standard rubric for all the papers sums the situation up by including "final answers should be given to a degree of accuracy appropriate to the context". Three significant figures may often be the norm for this, but this always needs to be considered in the context of the problem in hand. For example, in quoting from Normal tables, *some* evidence of interpolation is generally expected and so quotation to four decimal places will often be appropriate. But even this does not always apply – quotations of the standard critical points for significance tests such as 1.96, 1.645, 2.576 (maybe even 2.58 – but not 2.57) will commonly suffice.

Talking now in general terms, the examiners always exercise sensible discretion in cases of small variations in the degree of accuracy to which an answer is given. For example, if 3 significant figures are expected (either because of an explicit instruction or because the general context of a problem demands it) but only 2 are given, a candidate is likely to lose an Accuracy mark; but if 4 significant figures are given, there would normally be no penalty. Likewise, answers which are slightly deviant from what is expected in a very minor manner are not penalised (for example, a Normal probability given, after an attempt at interpolation, as 0.6418 whereas 0.6417 was expected). However, there are increasing numbers of cases where candidates give answers which are *grossly* over- or under-specified, such as insistence that the value of a test statistic is (say) 2.128888446667 merely because that is the value that happens to come off the candidate's calculator. **Such gross over-specification indicates a lack of appreciation of the nature of statistical work and, with effect from the January 2011 examinations, will be penalised by withholding of associated Accuracy marks.**

Candidates must however always be aware of the dangers of premature rounding if there are several steps in a calculation. If, say, a final answer is desired that is correct to 3 decimal places, this can in no way be guaranteed if only 3 decimal places are used in intermediate steps; indeed, it may not be safe to carry out the intermediate work even to 4 decimal places. The issue of over-specification may arise for the final answer but not for intermediate stages of the working.

It is worth repeating that most candidates act sensibly in all these respects, but it is hoped that this note may help those who are perhaps a little less confident in how to proceed.

G241 Statistics 1

General comments
The level of difficulty of the paper appeared to be entirely appropriate for the candidates with a good range of marks obtained. Question 4 proved to be the most challenging question on the paper and question 7 the easiest. Very low scores were rare and very few candidates seemed totally unprepared. There were, on the other hand, a good number of almost completely or completely correct scripts. There seemed to be no trouble in completing the paper within the time allowed.

Most candidates supported their numerical answers with appropriate explanations and working although some rounding errors were noted particularly in questions 5 and 6. Arithmetic accuracy was generally good although there is still evidence of candidates not being proficient or sensible in their use of calculators. In particular the simplest method of doing question 5(i) is by use of the statistical functions on a calculator, but few candidates used this approach. Amongst some candidates, there was evidence of incorrect use of point probabilities instead of tail probabilities in question 6 and of a totally wrong method to establish outliers in question 1.

The scripts were invariably well presented and legible with the use of a pre-printed answer book not appearing to constrict candidates' work; most candidates were able to answer in the space provided in the answer book, and only a few used additional sheets.

Comments on individual questions

1) (i) Most candidates gave the correct answer of positive skewness although a few thought that the skewness was negative; the occasional response of 'skewed to the right' was not acceptable.

1) (ii) The answer of 2.3 for the IQR was obtained by most candidates. Wrong answers included (17.6 – 7)/4 = 2.65 and errors in stating the value of the upper quartile. Many candidates made mistakes in finding the boundaries for outliers with the use of median ± 1.5 × IQR being very common. Those who used the quartiles occasionally combined the values with multiples of 1, 2 or even 3 of the IQR. The use of the limits to establish the presence of outliers, or otherwise, was good although a number of candidates used a value of 18 rather than 17.6. This error was treated generously. Some candidates curiously tried to argue in terms of the standard deviation.

1) (iii) There were many sensible and complete answers with the most common including 'an error in the data', 'no lower outlier due to the minimum wage' and 'the outlier being a manager or supervisor'. Some candidates only gave one reason or just concentrated on one end of the data. A very few candidates just repeated the information about outliers given in part (ii).

2) (i) This was very often correct but a number of candidates stopped when they had worked out the first two terms. Some candidates tried to sum the terms without k or the k became an afterthought after the summation was completed.

2) (ii) The calculation of E(X) and Var(X) was well executed on the whole. There were still some candidates who mistakenly divided E(X) and/or Var(X) by 4. Some forgot to square E(X) in the calculation of the variance. A few candidates thought that they could attempt this question without using any probabilities.

3) (i) Most candidates knew that they had to find frequency density and on the whole were very successful. Occasionally-seen errors were attempts to multiply frequency by width or divide by mid-interval or divide width by frequency. However the most common error was to use the given frequencies as the heights. Labelling was not always successful, and although a pleasing number of candidates knew that the label should be frequency density, some gave it simply as frequency, and those using a non-unitary class width as standard often had difficulty indicating this correctly on the graph. The vertical linear scale was usually correct (and sensible!). On the horizontal scale the majority of candidates were able to get the width of the bars correct, but a number of candidates thought that they should number their scale with inequalities rather than giving a correct linear scale. Very few candidates mistakenly left gaps between the bars. Use of rulers is to be encouraged to produce a clearer diagram.

3) (ii) This was usually well done although some candidates seemed to think the frequency was 100. Although candidates should have been trying to find the 45.5th value many were looking for the 45th value; this error was not penalised. Many candidates failed to indicate that it was the 45th value not just 45 that was in the correct interval.

4) (i) This was the least well done question on the whole paper. The majority of candidates had 0.2 × 0.2 in some form as their answer.

4) (ii) This part was done slightly better, but even so the correct answer of 0.0384 was fairly unusual. Common errors included $({}^{1}/_{5})^{5}$ and ${}^{1}/_{5!}$ whilst some candidates tried to use some form of binomial probability.

4) (iii) Most candidates managed to subtract their answer to part (ii) from 1 although some made arithmetical errors whilst others did not attempt this part at all.

5) (i) This question was answered more successfully than in the past. There were many wholly correct solutions, usually showing full working but occasionally by use of calculator. The vast majority of candidates found the mean correctly, although a number of incorrect answers were seen including ${}^{75}/_{59}$, ${}^{59}/_{6}$ or ${}^{59}/_{5}$. Some lost a mark because of inappropriate rounding of their answer. Many candidates found the standard deviation correctly but there was a wide variety of wrong methods including finding $(fx)^2$ or xf^2 instead of fx^2. A few candidates correctly found fx^2 but then forgot to subtract $(\sum x)^2/n$ or used 59, 58 or 6 rather than 75 as the value of n. Only a few candidates divided by 75 and thus found the root mean square deviation and only a few forgot to square root their variance. Candidates who obtained ridiculously large answers often did not seem to realise that their answers could not possibly be correct.

5) (ii) Most candidates found the new mean successfully. However many stated that the standard deviation would not change. Units were often missing or only given for the mean. A number of candidates gave the new mean as 0.82p rather than 82 pence or £0.82. Some candidates did not realise that they could just multiply their answers to part (i) by 1.04 and instead multiplied the numbers of loaves by 1.04 then recalculated the new mean and standard deviation.

6) (i)(A) The vast majority of candidates found the correct value of 0.2385, with most preferring to use a binomial expression rather than tables. Occasionally an answer of $P(X \leq 2) = 0.7338$ was seen.

6) (i)(B) Candidates were less successful in this part, with mistakes occurring due to rounding errors when using the point probability approach, the omission of a term such as P(X = 0) when using point probabilities, misuse of tables, or answers such as 1 – 0.2835, 1 – 0.9018 or 1 – 0.4503 rather than 1 – 0.7338.

6) (i)(C) This was very well answered although a significant number of candidates rounded to 2 or even 1, losing a mark.

6 (ii) Many candidates constructed the hypotheses correctly although a few used "equals" for H_1. The main loss of marks came from poor notation such as P(X) = 0.1, H_0 = 0.1, X=0.1, P(0.1), etc. However many candidates still failed to define p as the probability that a randomly selected tile is faulty. Virtually all adequately explained why H_1 took the form it did.

6) (iii) Few candidates confidently scored full marks in this part. Some candidates often had little idea as to where to begin; other candidates used point probabilities and even those who used the correct probabilities of 0.0982 and 0.0282 with a comparison of 0.05 often started the critical region at 4 rather than 5. Occasionally the critical region was given as 'from 5 to 8' rather than 'from 5 to 18'. Some candidates failed to show necessary working; an answer along the lines of 'the first value in tables above 0.95 is 4, so critical region is 5 to 18' did not score full marks. A very small number of candidates thought their comparison should be with 0.025 rather than 0.05. As has been stressed in past examiners' reports, candidates must quote specific probabilities in finding critical regions and then explicitly compare these probabilities to the significance level. If they do not do this they may not get any marks.

Although it is given in the mark scheme, it is worth repeating here the recommended method for comparing the probabilities with the significance level. Candidates should find the two upper tail (in this case) cumulative probabilities which straddle the significance level.

$P(X \geq 4) = 1 - P(X \leq 3) = 1 - 0.9018$ or $0.0982 > 5\%$

$P(X \geq 5) = 1 - P(X \leq 4) = 1 - 0.9718$ or $0.0282 < 5\%$

6) (iv) This was poorly answered with very few using their critical region and stating that 4 was outside it. Most successful answers started again with 0.0982>0.05; often those using this approach wrote down 0.0282<0.05 and wrongly rejected H_0. The other major error was to finish by saying 'insufficient evidence to reject H_0' and then making no reference to the context. The use of point probabilities was again frequent, even by some candidates who had successfully used cumulative probabilities in part (iii).

7) (i) This question provoked a variety of responses. Those who read the question carefully often gained high marks but there were many candidates who got off on the wrong foot by either having too many sets of branches on their tree diagram e.g. (0900, 1000, 1100, 1200) or having the 1000 branches labelled with probabilities of 0.5 instead of the correct 0.95 and 0.05. Some candidates omitted some or all of the labels for 'on time' and 'late'.

7) (ii)(A) Whether using their tree diagram or not, this was well answered and most candidates gained both of the marks.

7) (ii)(B) Most candidates were able to trace their way through the tree diagram to achieve the correct response of 0.0325. A generous follow through was in place for those who may have made an error in one or more of their probabilities.

7) (ii)(C) This was again well answered by most candidates, and once again a generous follow through was in place. Candidates should be reminded that total accuracy in intermediate working is important when dealing with probabilities e.g. 0.857375 should not be rounded to 0.86. The cumulative effect of 4 prematurely rounded probabilities caused some candidates to have their final answer outside the required range.

7) (iii) Many candidates realised they had to evaluate the 1000, 1100 and 1200 on time and the 1000 on time, 1100 late and 1200 on time to reach 0.885875 or its equivalent on follow through. Often they stopped at this point, not realising the conditional probability requirement of the question. The more discerning candidates used their answer from part (ii) (C) to complete the question successfully.

G242 Statistics 2

General comments

This year saw another small entry, similar in size to last year. The majority of this year's candidates were very well prepared and many high marks were produced.

Overall, the candidates demonstrated very good understanding of the statistical methods required and communicated their responses using appropriate statistical terms and in sufficient detail. The parts of questions requiring candidates to interpret information, explain or comment were not as well answered as the parts involving calculation. Some candidates lost marks through incorrect use of their calculator; there were several cases where a correct method was seen but the final answer did not match what was written. Problems identifying the correct number of degrees of freedom were again common.

Comments on individual questions

1) *(Chi-squared test for Association)*

(i) Some candidates mixed up the hypotheses, leading to contradictory conclusions and loss of marks. Some candidates did not include context in either their hypotheses or in their concluding remarks. A few slips with degrees of freedom were seen and incorrect critical values were fairly common. It is expected that candidates should state the number of degrees of freedom used – some did not and were penalised.

(ii) This part was poorly understood. Few candidates showed an understanding of the link between the size of the contribution to the test statistic and the level of association. For willow warblers and chiffchaffs, candidates were expected to identify the cells containing relatively large contributions and comment whether this provided evidence that the warblers were seen more frequently or less frequently than expected in the corresponding tree. For whitethroats, the candidates were expected to comment that the small contributions indicated that they occurred in numbers that would be expected if there were no association between warbler and type of tree.

(iii) This too was poorly answered. Generally, candidates interpreted the question incorrectly, not realising the importance of the condition that the bird heard was a whitethroat. A small number reversed the question, finding the probability that the bird was a whitethroat given that it was singing from a birch tree.

2) *(Hypothesis test using the t distribution)*

(i) This required an understanding of the differences between the situations leading to hypothesis tests based on the Normal distribution and the t distribution. In general, this was not well answered. Several candidates did not comment on the assumption necessary for a t test to be valid. Confusion between population and sample was evident.

(ii) This required candidates to provide estimates for population mean and population standard deviation. This led to full marks in most cases.

(iii) This part was well answered. Occasional marks were lost for failing to define μ as the population mean or for writing hypotheses in terms of some other variable (e.g. x). Candidates were expected to give hypotheses in terms of μ rather than in words. Several candidates stated a correct calculation for the test statistic but did not calculate it correctly; a final value of 0.15075 was seen more than once. Some candidates were unsure of the value to use for the number of degrees of freedom. Some stated a 2-tail critical value despite intending to use a 1-tailed test. Conclusions were stated in appropriate terms, were not too assertive and were given in the context of the question.

3) *(Chi-squared test for goodness of fit)*

(i) (*A*) was well answered. Most candidates successfully verified the sample mean as 2.4 using the given frequency distribution.
(*B*) was poorly answered with several candidates making comments about results being random and/or independent, rather than comparing the mean and variance. Several compared mean and standard deviation and earned no credit.

(ii) In general, candidates could find P(X = 1) but many struggled with P($X \geq 6$).
The remainder of the question was well answered – some lost marks by using a critical value from the t distribution and some associated 'significant' with not rejecting the null hypothesis.

4) (*Wilcoxon test)*

This was generally well answered. Candidates were required to provide values in their hypotheses and make it clear that the values referred to the population median. Many lost marks here. In the remainder of the question, marks were lost for providing an incorrect critical value. Some contradictory conclusions were seen (as in Q3 (ii)).

5) *(Use of Normal distribution (confidence interval))*

(i) This was well done.

(ii) This was less well handled, with +2.326 leading to an answer of 497 seen on several occasions. Candidates should be encouraged to sketch diagrams to help ensure sensible answers are found.

(iii) This was well answered although some used 1.645 in place of 1.96.

(iv) In this part, many were not convincing in their explanations. Candidates were expected to point out that the value of 9 (gallons) was 'below' the confidence interval – stating that it was 'not contained in' the confidence interval was deemed not to support the statement that the mean value was less than 9 gallons.

G243 Statistics 3

Comments on individual questions

1) (i) Most candidates gained full marks here, but some had axes starting at for example 90 without using a broken scale. Others did not know how to draw a scatter diagram, instead plotting the runners' 'names' on the *x*-axis and the times before and after on the *y*-axis.

1) (ii) Very few candidates used their calculators' built in functions to calculate the correlation coefficient and there were a number of errors in using the formula to calculate the coefficient.

1) (iii) Most candidates gave their hypotheses correctly in terms of ρ, but hardly any defined ρ. Most found the critical value although a few gave the two-tailed value instead. Most candidates then completed the test correctly. A very few tried to carry out other types of test.

1) (iv) Most candidates correctly said that a bivariate Normal distribution was required and most knew that an elliptical scatter diagram would indicate that the sample came from this distribution.

1) (v) Many correct answers were seen but a number of candidates thought that the null hypothesis was 'no correlation' rather than the correct 'no association'.

2) (i) There were a number of fully correct solutions. Almost all candidates found the mean and most also found the variance or standard deviation correctly. The majority of candidates made a good attempt at the test statistic, but occasionally there were errors such as using s instead of s^2 in calculating the pooled variance. The hypotheses were often correctly given either in symbols or, more often, in words, but most candidates failed to mention 'population' when giving hypotheses in words. The critical value was usually correct, and most candidates who had got this far went on to complete the test correctly. A very small number of candidates tried to carry out a *t*-test.

2) (ii) Most candidates said that the samples were large, but rather less went on to explain that the sample variances could be used as if they were the population variances without appreciable error as the samples were large.

2) (iii) The majority of candidates correctly realised that the lower variance of the carrots from Supplier B might make them preferable.

3) (i) Most candidates ranked the data and found the Wilcoxon test statistic, although a few attempted to carry out a paired test despite the values of *n* being different. Most then went on to complete the test correctly, although a few compared the Wilcoxon test statistic to the Mann-Whitney critical value (it would of course be acceptable to carry out the whole procedure according to the Mann-Whitney formulation). Most gave their hypotheses in terms of the median, although a few incorrectly wrote average or even mean. Many failed to mention population.

3) (ii) Most candidates answered this well although several did not mention that the test should be unpaired.

3) (iii) This was not well answered. Very few gave a sensible reason for a signed rank test being inappropriate.

4) (i) Many candidates correctly identified the population as consisting of the concrete, but few gave a fully correct description – 'all of the concrete in all of the lorries'.

4) (ii) Most candidates realised that the third method was the best and correctly identified the failings in the first two methods.

4) (iii) There were many correct solutions to this part, although several failed to mention repeats. A few tried to describe a systematic sample.

4) (iv) Some candidates gave the hypotheses correctly in symbols and defined their symbols. Rather more gave them in words, usually correctly. Few stated the necessary assumption of Normality of both populations. Most candidates found the pooled variance although several candidates gave their result as 0.2473, the standard deviation. The method for finding the test statistic was well known and candidates could follow through to gain the final marks (provided that they scored the method marks here). Most realised that they needed to use 13 degrees of freedom and found the correct critical value, usually going on to complete the test correctly.

4) (v) There was a wide variety of wrong answers here, such as 'the samples are small', or 'there are equal numbers in the two samples'. Very few gave an answer that was reasonable.

4) (vi) Most candidates gained full marks here, although a few had the wrong critical value and scored zero.

OCR (Oxford Cambridge and RSA Examinations)
1 Hills Road
Cambridge
CB1 2EU

OCR Customer Contact Centre

14 – 19 Qualifications (General)
Telephone: 01223 553998
Facsimile: 01223 552627
Email: general.qualifications@ocr.org.uk

www.ocr.org.uk

For staff training purposes and as part of our quality assurance programme your call may be recorded or monitored

Oxford Cambridge and RSA Examinations
is a Company Limited by Guarantee
Registered in England
Registered Office; 1 Hills Road, Cambridge, CB1 2EU
Registered Company Number: 3484466
OCR is an exempt Charity

OCR (Oxford Cambridge and RSA Examinations)
Head office
Telephone: 01223 552552
Facsimile: 01223 552553