RECOGNISING ACHIEVEMENT

GCE

Statistics (MEI)

Advanced Subsidiary GCE **AS H132**

Report on the Units

June 2008

H132/MS/R/08

OCR (Oxford, Cambridge and RSA Examinations) is a unitary awarding body, established by the University of Cambridge Local Examinations Syndicate and the RSA Examinations Board in January 1998. OCR provides a full range of GCSE, A level, GNVQ, Key Skills and other qualifications for schools and colleges in the United Kingdom, including those previously provided by MEG and OCEAC. It is also responsible for developing new syllabuses to meet national requirements and the needs of students and teachers.

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 syllabus 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

G241 Statistics 1

General Comments

The standard this summer was variable. There were some excellent scripts seen by the examiners reflecting the hard work and dedication of teachers, lecturers and candidates. On the other hand there were a substantial number of candidates who seemed totally out of their depth who struggled to make any real progress.

Candidates should be reminded to work with total accuracy and not to round their answers severely as they progress through a calculation.

It was pleasing to see that a number of centres had acted on comments made in previous reports particularly with regard to the definition of p in the construction of hypotheses.

Comments on Individual Questions

1) The calculation of an estimate of the mean and standard deviation of grouped data presented unexpected problems for a sizeable number of candidates. Often 2 or more mid-points of the classes were incorrect thus throwing out any possibility of achieving the accuracy required. A common error even by the better candidates was to use mid-points of 1.5, 4.5, 6.5, 8.5 and 15. Some candidates had little idea how to obtain the mid-points and thought that the mean could be somehow calculated from multiplying the frequencies by the class widths or the frequencies by one of the boundary values. It was disturbing to see many candidates attempting to work out the standard deviation without using any frequencies. This is clearly a topic which deserves more attention to precision and process for the future.

   The concept of finding the upper boundary for any outliers was well known in terms of mean + 2 standard deviations but several tried to argue the case with Q_3 + 1.5 IQR (not that these data were available) or insisted using mean + 1.5 standard deviations. Candidates should be careful not to make rash statements such as 'there **are** outliers in the data' but instead be more circumspect and claim that 'there could be or may be some outliers in the final class'.

2) The work on testing for independent events was pleasing with a variety of methods used by candidates. Most went down the route of showing numerically that $P(W) \times P(C) \neq P(W \cap C)$ and hence the events were not independent. Some tried their luck with non numerical or qualitative attempts but to little avail.

   The Venn diagram was, unfortunately, often lacking in credibility. There are still too many candidates filling in the various regions with the incorrect probabilities. The region $W \cap C'$ was often given as 0.2 instead of the correct 0.14 and likewise the other region $C \cap W'$ was written as 0.17 instead of the correct 0.11. The region $W \cap C$ was invariably correct as 0.06. A curious number of candidates often labelled the region $W' \cap C'$ as 0.63 instead of the correct 0.69. Again, this is an area that deserves the attention of candidates for future examinations.

   The calculation of P (W/C) was well attempted and most scored 2 marks. The conclusion was usually sound but many did not choose their words carefully and quoted '**more** children speak Welsh' when really they meant 'the proportion of children speaking Welsh is higher.'

3) (i)(*A*) Many candidates had difficulty composing an equation which included p + q and a summation to 1.

(*B*) A little better, with some realising that the equation for E(X) must now include 2p + 3q.

(*C*) The solution of the resulting simultaneous equations seemed to be off the mathematical radar for many candidates with many struggling to find solutions for p and q.

(ii) The variance was usually calculated correctly bearing in mind that a generous follow through was applied for those candidates who did not find the exact values of p and q earlier. The only common error was the omission of 0.67^2 leaving an answer of 1.07.

4) This was a popular question which was well answered by many candidates. In (i) part (A) most gained the correct answer of 0.6634 but then did themselves no favours by unnecessarily rounding the answer to 0.66. Part (B) was well answered but there was some confusion about the meaning of P(X>1). Some believed it to be 1 – P(X=0) rather than the correct form of 1 – {P(X=0) + P(X=1)}. In the last part, most knew the E(X) = np formula and gained the marks, even on follow through.

5) Candidates need to be reminded that a hypothesis test on the binomial distribution requires an initial set up of the following conditions.

- The definition of the parameter p, in context
- The use of the correct notation for H_0 and H_1, namely in the case of this question that H_0: p = 0.35 and H_1: p > 0.35
- A clear explanation, in context, of why H_1 takes the form that it does.

Unfortunately, many omit the requirements of the first and last bullet points, thus losing 2 valuable marks. It is worth reminding centres again that sloppy or poor notation such as H_0: P(x = 0.35) and H_1: P(x > 0.35) is penalised by the examiners. Too many candidates are prone to this form of notation.

Many otherwise worthy initial set ups were spoilt by candidates using point probabilities or selecting the wrong tail. It was not uncommon to see $P(X \geq 8) = 0.0422$ when, in fact this was $P(X \geq 9)$. The correct solution required $P(X \geq 8) = 1 – P(X \leq 7) = 1 - 0.8868 = 0.1132$. Some candidates wrote ridiculous statements along the lines of 0.9578 > 5%. It must be emphasised, once again, that the tail probability must be compared with the significance level of the test. All further marks in the question are dependent on this important fact. The next stage is to accept or reject H_0 and then reach a valid conclusion in context.

6) There were many successful attempts to the first half of this question. Candidates were able to demonstrate a good understanding of probability calculations using their tree diagrams.

Part (A) was invariably correct as 0.04. Most were able to achieve 0.9559 in part (B) by adding the 5 separate probabilities but very few candidates realised the quick way to achieve the answer by $1 – 0.21^2$ =0.9559. A common error in part (B) was the omission of the 0.79^2 term giving 0.3318 as an answer.

In part (C) most candidates preferred to list and add the 4 probability terms to gain 0.9801. Relatively few spotted the quick way of 0.99^2 would reach the same answer. Some candidates made the error of believing that neither of the people was born overseas could be calculated from $1 – 0.01^2 = 0.9999$. The conditional probability in part (ii) elicited some very good responses with most realising the correct method although some did write (0.04 x 0.9801)/0.9801with depressing regularity.

Only the better candidates made any progress in part (iii) with many finding $1 – 0.79^5$. Some candidates had become muddled by this stage and it was not uncommon to see $1 – 0.21^5$ or even $1 – 0.9559^5$. The latter two methods did, however, attract a partial award. Part (iii) (B) was often well attempted by the better candidates with equally as many opting for using logarithms as for using a trial and improvement method.

7) Part (i) was almost invariably correct with the response of positive skewness.

Part (ii) was well tackled with many achieving the answer of 950 000 but some candidates left their answer as 950 and lost a mark.

Many reached the required cumulative frequency of 2150 (thousands) via 1810 + 340 but there were instances of 1810 + 345 seen by the examiners. Almost all candidates were able to locate the position of the median as the 1385 or 1385½ value. Only the very talented candidates were then able to carry out the linear interpolation of

$30 + \frac{145}{570} \times 10 = 32.54$, to achieve the median age.

It was pleasing to see many successful attempts at finding the frequency densities in part (iv). Without doubt, the frequency divided by class width was the most popular method but other strange but nevertheless correct methods were seen. The resulting histogram was well drawn but some candidates did make life difficult for themselves by choosing a bizarre scaling (e.g. 3cm = 10 units on the vertical axis).

The comments in part (iv) were often not what the examiners were looking for. Many opted to compare numbers across the two histograms but it should have been evident that **all** the populations for **each** age group were higher in Outer London than Inner London. Some candidates did pick up on the salient points of the two histograms by comparing the different modal classes (20 – 30 for Inner London; 30 -40 for Outer London). In making comparisons it is advisable that candidates mention proportions rather than refer to ‘more than’ or ‘less than’ statements.

Part (vi) elicited some positive responses with many realising that the mean, midrange and standard deviation would all increase in the light of the new information. Some thought the standard deviation would decrease rather than increase but most knew the interquartile range would be unchanged.

G242 Statistics 2

General Comments
In the third sitting of this AS Statistics module the size of entry has increased sufficiently to enable a more informative report.

This year a broad range of marks was seen. Many candidates displayed a good understanding of the range of statistical techniques required. It was pleasing to see that most candidates managed to choose appropriate techniques for each question. In the hypothesis tests, many candidates provided appropriate hypotheses, but only the better candidates successfully defined the symbols used. In hypothesis tests, candidates are expected to give conclusions using the context in which the problem is set, also, conclusions should not be too definite; candidates should use phrases such as “the evidence suggests that” rather than “this proves that” – this year’s candidates seem to have taken this advice on board. Understandably, candidates seem to be more comfortable applying a Chi-squared test for ‘association’ than they do for ‘goodness of fit’; to prepare for this exam, candidates should develop a good understanding of how to apply the various probability distributions, making use of the statistical tables provided, and familiarise themselves with the various conditions necessary to justify their choice of probability distribution or hypothesis test.

Comments on Individual Questions

1) **Poisson and Normal distribution calculations**

Some candidates scored well on this question. In part (i), candidates were expected to quote the conditions for a Poisson model to apply, giving reference to the context of the question. For example, merely stating ‘independent’ or ‘independence’ was not sufficient to gain credit – pleasingly, most candidates realise this and attempt to put their comments in context. Part (ii) was well answered, with candidates making good use of tables in (ii) *B*. As expected, part (iii) was answered successfully by only the better candidates. Candidates should watch out for ‘binomial’ situations to arise following routine probability questions. Part (iv) was well answered; in general, candidates should be aware that when an answer is provided, sufficient evidence of working must be seen. In part (v) the better candidates managed to identify the correct z value and ‘de-standardise’ accordingly.

2) ***t* test**

Part (i) was well answered. In part (ii), the test, candidates were expected to state their hypotheses in terms of the population mean, μ, and also define μ as the mean decrease in cholesterol level for the underlying population. Many candidates struggled to do this successfully. Many candidates managed to obtain the correct test statistic and proceed to make a sensible comparison with the correct critical value. To gain full marks for their conclusions, candidates should again refer to the context of the question and not be too assertive in their comments, as outlined in the general comments above. In part (iii) many candidates showed that they were familiar with the necessary assumptions.

3) **Chi-squared test for Association**

Parts (i) and (ii) of this question were well answered. Most candidates manage to state suitable hypotheses, obtain expected frequencies and calculate a value for the test statistic. Candidates should state the number of degrees of freedom in addition to the critical value. In such tests, providing a test statistic based on a two-tailed test, as was seen in this sitting, is not looked upon favourably. Part (iii) proved difficult, although some candidates provided acceptable interpretations.

4) **Wilcoxon test**

Overall, this question was handled quite well, most candidates appearing to be quite familiar with the process. However, some failed to use ranks at all, and were heavily penalised. Those using ranks tended to go on to a successful completion of the test, but some slips were made when calculating the difference between the sample values and the median provided by the null hypothesis. In this type of question it is not unusual for candidates to rank the differences rather than their absolute values – this was not seen this time.

5) **Chi-Squared test for goodness of fit**

This question was reasonably well answered. In part (ii), candidates were expected to use the cumulative probability tables to obtain probabilities and expected frequencies; better candidates did so quite competently. Mistakes were made in calculating $P(X \geq 6)$, with $P(X = 6)$ seen. In carrying out the test, only the better candidates realised the need to merge the cells for $X = 5$ and $X \geq 6$; marks were still available for those neglecting to do this.

G243 Statistics 3

General Comments

This was the third sitting of this module and the third occasion on which the new award of AS Statistics was available. Though the entry was small – 31 candidates from 6 centres – it was a considerable increase over the previous two years. It is to be hoped that this indicates that the AS Statistics specification is now beginning to become established. Users and potential users of the specification are reminded that there is now an MEI textbook explicitly supporting it, and this will no doubt be helpful. Other textbooks may of course be useful too.

The specification and the question papers and mark schemes are published separately in the normal way. Teachers are warmly invited to study these and consider using this qualification as a support for the very many subjects where statistics is used as a tool. This does not exclude using the qualification alongside A/AS-level mathematics (though the first module, which is common to this specification and to the MEI A/AS mathematics suite, naturally cannot be counted towards an award in both).

The question paper consists of four questions, with question 4 being longer than the others, following through a more extended line of enquiry with way-stages along the route. All questions are to be attempted.

There was some good work this year, candidates showing understanding of the range of techniques covered in this module and, arguably more important, a good grasp of the statistical principles underlying them. Sadly there was also some very poor work, candidates showing virtually no understanding of any of these ideas. And, as might be expected, there was work of an intermediate standard, candidates doing reasonably well in some areas but less so in others. This variety in the quality of candidates' responses is not dissimilar to what normally occurs with any mathematics-based module, and is perhaps another indication that the specification is settling down and becoming established – though it is earnestly to be hoped that the very poor work will, at the least, become less poor in future years.

It is worth repeating two general points that were made in last year's report, for these faults continued to appear this year. First, some candidates were too assertive in stating conclusions of statistical tests. No statistical test can prove that any hypothesis is right or wrong, no matter what level of significance is attained. It is correct to state that "the result is significant [or not significant, as the case may be] at [say] the 5% level", and candidates are indeed normally expected to say that. But a subsequent verbal conclusion in the context of the problem is usually expected, and this should contain phrases such as "there is evidence that ..." or "it seems that ..."; it is not correct to simply state, assertively, "and therefore the means are not equal" [or whatever the context is]. Secondly, many candidates were not sufficiently careful in distinguishing populations and samples; this is a key distinction, and recognising the importance of it is one of the central tenets of the specification.

A third general point may be made. This refers to the selection of statistical tests to be used in dealing with the problem that is set out in a question. Correct selection of tests is another central tenet of the specification. In some questions there will be guidance, which might sometimes be fully explicit, but there will also be questions in which it is for the candidate to decide (for instance, whether a procedure based on N(0, 1) or on a *t* distribution is to be used). Incorrect selection in these circumstances is likely to lead to quite heavy loss of marks.

These general criticisms should not hide the fact, as stated above, that there was some good work this year. Clearly there are candidates who have benefited from this specification and understand it well.

Comments on Individual Questions

1) In part (i) of this question, candidates were first expected to identify the sampling scheme that had been described as quota sampling and then indicate an advantage and a disadvantage of it. Though some candidates were correct in this identification, many other suggestions were made. The "nearest miss", though still a miss by some distance, was perhaps stratified sampling, but there were several other suggestions. Some credit was given for advantages and disadvantages where the work generally followed-through in a reasonably sensible way.

Part (ii) started with a request for a scatter diagram and then asked for calculation of Spearman's rank correlation coefficient followed by a test based on it. Mostly the scatter diagrams were satisfactory, though it was extremely disappointing to find that some candidates could not even do this properly. Likewise there were some candidates who clearly had no idea at all how to do the test, but those who did know it usually did it satisfactorily. A fairly common error was to overlook the "one minus" component at the start of the familiar formula, leading to a result of 0.4391 instead of 0.5609. There were also some candidates who were wildly incorrect; they should know that a value of a correlation coefficient outside the range [–1, 1] *must* be wrong, and they should go back and check their work.

The question ended with a request for a short discussion as to whether the product moment correlation coefficient should have been used. Comparatively few candidates gave the correct answer in terms of an underlying bivariate Normal distribution and whether the scatter diagram suggested that such an assumption is reasonable (it isn't!). Some candidates talked about "outliers", for which some credit was given as it does impinge on the general idea. Many candidates talked about "linearity" of the plot, which does not really bring out the point.

2) This was on a test using N(0, 1) for comparing two means, given large samples.

Candidates were first asked to state the null and alternative hypotheses. This was often disappointing; some answers were fully correct, but there were many where not quite sufficient care had been taken. The general point made above concerning distinction of populations and samples is important here. Simply stating a hypothesis that "the means are equal" is inadequate. The *sample* means are clearly not equal, so there is nothing to test in their regard; what we want is to use them to reflect on whether the *population* means may be assumed equal. The use of the customary statistical notation in terms of μ for a population mean is efficient and likely to lead to correct statements of hypotheses, and is to be commended. Nevertheless, it is entirely permissible for candidates to state their hypotheses in words without any symbols at all, but there is an absolute insistence that the word "population" appears as appropriate if this is done.

The next steps were to calculate the sample means and sample variances, use them in the usual expression for the test statistic, and draw conclusions in the usual way by referring the value of it to the N(0, 1) distribution. Mostly this was done satisfactorily, except by the very poor candidates mentioned in the general comments, who usually had almost no idea what to do. Another exception, more widespread, was that the comparison with N(0, 1) was quite often done incorrectly. Some very strange incorrect methods appeared.

Finally, candidates tended to know that the lack of need for an assumption of underlying Normality had something to do with the largeness of the samples, though the explanations were not always fully complete or convincing.

3) This question was on the Wilcoxon signed rank test.

Apart again from the very poor candidates, most showed a general understanding of the basics of the procedure. However, several candidates made a serious error when setting off, in that they did not *rank* the differences, trying instead to use the procedure based on the differences themselves. Another fairly common error was to get the comparison with the tables the wrong way round: the observed value of the test statistic (23) is *less* extreme than the tabulated critical point (8) and so the result is *not* significant.

The final two parts of the question invited discussion about the "design" aspects of the statistical procedure. This is seen as an important aspect of the AS Statistics specification – candidates should understand *why* a particular procedure should be used and *how* it is to be implemented, as well as being able to undertake the calculations. The discussions were generally quite sensible but perhaps without full exploration. As an illustration, in part (iii) some credit was given for mentioning a factor that might affect both airlines (eg time for collecting luggage) but more for discussing that this might affect one airline to a greater extent than the other. [As an aside, one or two candidates talked about flights being delayed by accidents – not severe ones, the examiner hopes!]

4) This question opened with some "design" discussion, about the need for randomisation and replication. Many candidates had reasonable ideas here. It was quite a common answer that replication allowed results in some sense to be "averaged"; there is some sense in this (and some credit was duly given), but it does not quite capture the point that the inherent variability can actually be measured.

Part (iii) required an unpaired *t* test. Here there were several cases of incorrect test selection; it was particularly bizarre that some candidates tried to undertake a paired test with samples of unequal size! Assuming the correct procedure was selected, an important first step is to form the usual "pooled estimate" of the assumed common underlying variance (an assumption that should have been stated, along with Normality of *both* underlying populations, earlier but often wasn't). Most candidates knew how to do this, but some did not. Moving on to the test statistic, an occasional error was to omit to take a square root in the denominator. Further errors occurred with reference to the wrong *t* distribution (i.e. an incorrect number of degrees of freedom) and, even if this was correct, selection of the wrong critical point. Nevertheless, there were some candidates who did this part well and scored highly. But there were rather too many who went quite badly wrong somewhere.

In part (iv), candidates usually had a reasonable idea of the purpose of the pairing.

In part (v), the value of a test statistic for a further test was given. Candidates had to carry out the test, with appropriate discussion. Again there was a quite common error of the wrong number of degrees of freedom for a *t* distribution, and again wrong critical points were selected. For the required distributional assumption, most candidates knew that this had something to do with underlying Normality, but very few made reference to the underlying population of the differences.

Grade Thresholds

Advanced Subsidiary GCE Statistics MEI (H132)
June 2008 Examination Series

Unit Threshold Marks

Unit		**Maximum Mark**	**a**	**b**	**c**	**d**	**e**	**u**
G241	Raw	72	53	45	38	31	24	0
	UMS	100	80	70	60	50	40	0
G242	Raw	72	56	49	42	35	28	0
	UMS	100	80	70	60	50	40	0
G243	Raw	72	56	48	40	33	26	0
	UMS	100	80	70	60	50	40	0

Specification Aggregation Results

Overall threshold marks in UMS (ie after conversion of raw marks to uniform marks)

	Maximum Mark	**A**	**B**	**C**	**D**	**E**	**U**
H132	300	240	210	180	150	120	0

The cumulative percentage of candidates awarded each grade was as follows:

	A	**B**	**C**	**D**	**E**	**U**	**Total Number of Candidates**
H132	9.7	12.9	35.5	51.6	64.5	100	31

For a description of how UMS marks are calculated see:
http://www.ocr.org.uk/learners/ums_results.html

Statistics are correct at the time of publication.

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