edexcel

GCSE

Edexcel GCSE

Statistics 1389

Summer 2005

advancing learning, changing lives

Examiners' Report

Edexcel GCSE
Statistics 1389

Edexcel is one of the leading examining and awarding bodies in the UK and throughout the world. We provide a wide range of qualifications including academic, vocational, occupational and specific programmes for employers.

Through a network of UK and overseas offices, Edexcel's centres receive the support they need to help them deliver their education and training programmes to learners.

For further information please call our Customer Services on 0870 240 9800, or visit our website at www.edexcel.org.uk.

Summer 2005

Publications Code UG017075

Contents

1 PRINCIPAL EXAMINERS REPORT – PAPER 1F (FOUNDATION)

1.1 GENERAL COMMENTS

1.1.1 This paper was accessible to the majority of candidates and there was little evidence to suggest that they were short of time.

1.1.2 The average attainment of candidates this year was higher than last year. Even the weak candidates were able to score some marks on the higher grade question.

1.1.3 It was encouraging to see so many candidates bring a protractor to this examination. This ensured that the pie chart question was answered well by virtually everyone.

1.1.4 The presentation of work was generally good, but candidates should be reminded to use blue or black ink (not pencil) at all times- including diagrams. Candidates should also be advised to do all their calculations and checks within the space provided for the question.

1.2 REPORT ON INDIVIDUAL QUESTIONS

1.2.1 Question 1

This question was done well by virtually all the candidates. Errors were rare and usually in part (c)- commonly $4\frac{3}{4}$ or $1\frac{3}{4}$.

1.2.2 Question 2

This question was generally done well. In part (a), B was usually placed correctly, but a significant minority placed C to the right of B. In part (b), most candidates stated that an even number should be taken. Other correct answers included “greater than 5”, or “less than 6”. Some of the weaker candidates either confused the figures by stating “less than 5”, or gave an example which did not include the counters in the question.

1.2.3 Question 3

There were many confused answers in part (a). Rather than writing down the advantages of taking a sample, candidates often commented on the advantages of taking a pilot survey. Avoiding bias or inaccurate answers were common explanations. In part (b), "draw the names from a hat" was a very popular answer, but there were many who gave answers relating to systematic sampling. Some candidates failed to earn the mark if their answer described using random numbers (eg from a calculator) but did not mention that the students should first be numbered.

1.2.4 Question 4

Part (a) was answered well by the most of the candidates. In part (b), the vast majority of the candidates could not explain how to take a stratified sample. Few realised that it was necessary to take 9 black cows and1 brown cow from the herd, and if they did, they did not say that this should be done randomly. Common incorrect answers were "take 5 black cows and 10 of 5 brown cows" and "choose the cows".

1.2.5 Question 5

The question was answered well by the majority of the candidates. Most knew how to draw the composite bar chart and give a suitable key. The most common misunderstanding resulted in sections being drawn either as overlapping bars, or as small bars side by side- these answers were usually awarded a mark for an appropriate key. Virtually all the candidates answered part (b) correctly.

1.2.6 Question 6

This question was usually answered well. In part (a), the most common error was in confusing the mode, median and mean. Some of the weaker candidates did not order the numbers when finding the mode. In part (b), the median was sometimes given as a range or as two numbers when, perhaps, candidates forgot that the midpoint of the two numbers gave the median.

1.2.7 Question 7

In part (a), most confusion arose out of candidates thinking of the Census, rather than a census. Many gave spurious answers which were mere speculation and not related to the question. In part (b), candidates often failed to realise that taking a sample near to the proposed ring road, or taking a sample of motorists, would cause bias. In part (c), few candidates gained the mark as they were more concerned about the ease of answering closed questions than by the analysis of the results. Part (d), was answered relatively well.
Most candidates could identify the bias in the question, but many grappled with how this should be described. In part (d)(ii), candidates often gave a question which was just as biased as the question they were trying to replace. Only a minority considered the inclusion of response boxes.

1.2 .8 Question 8

Part (a)(i) was well answered. In part (a)(ii), a significant minority of candidates gave their answer out of 60 rather than 100. In part (b), most candidates gained a mark for stating that it did reduce the foot rot, but many failed to earn the second mark as they referred to the numbers getting foot rot rather than to the proportions.
In part (c), it was not clear whether many of the candidates had a clear idea of why the vaccine was not given to all the cows.
Some suggested it might be because of the possible side-effects, or that it was too expensive to give to all the cows. Many referred to

the need to give it to just one group of sheep, but did not earn the mark as they did not mention some comparison with the other group.

1.2.9 Question 9

This question was done well by the majority of the candidates. In parts (a) and (b), candidates were generally able to plot the points accurately and draw an appropriate line of best fit. Only a very small minority drew the line away from the points. In part (c), nearly all the candidates were able to use their line of best fit to find an estimate for the amount spent by customers. The most common error was to give the answer as2500 rather than 25000. In part (d), some candidates referred incorrectly to a point on the graph or to people's attitudes to advertising.

1.2.10 Question 10

Parts (a), (b) and (c) were answered well by the majority of the candidates. A common error in part (a) was to work out the IQR rather than the range. In, contrast parts (d) and(c) were answered poorly. In part (d), many candidates were unable to make an appropriate comparison between any specific features of the box plots (such as the medians). There were many vague answers such as "it is further up the scale", or "it is higher". In part (e), many candidates were able to score a mark for identifying the distribution with the greater skew. This was fortuitous in some cases as it was rare to see a correct description of the skewness.

1.2.11 Question 11

This question was answered well by the majority of the candidates. In parts (b) and (c), some candidates gave rounded answers. In Part (d), some candidates misread the question and suggested that the change in life expectancy was due to "better medicines".

1.2.12 Question 12

In part (a), the majority of candidates were able to complete the table with cumulative frequencies, but there were a considerable proportion who accumulated the upper class boundaries. In part (b), only the best candidates were able to draw a fully correct cumulative frequency diagram. Many candidates failed to use linear axes or struggled with labelling- it was not uncommon to see $\leq 5, \leq 10,$ etc. on the horizontal axis. Some, having plotted the points accurately, did not join them with either line segments or a curve. Hardly any candidates gained he mark for a title. In part (c), it was surprising to find that so few candidates knew how to read off the median. Many showed their confusion by reading off a value from half way up the vertical grid, or from half way along the *horizontal* scale.

1.2.13 Question 13

This question was done well by the majority of candidates. In part (b), most candidates referred to a division by 6 (or 60). In part (c), pie charts were generally part (e), many failed to relate the percentages for Sussex with the angles for Town Clinic. Some candidates only related one percentage to one angle and thought this was sufficient justification for their claim.

1.2.14 Question 14

Parts (a), (b) and (c) were answered well by most candidates, but only the best were able to tackle parts (d) and (e). Strangely some candidates, having given an incorrect answer in part (a), gave $\frac{5}{8}$ in the tree diagram in part (c). In parts (d) and (e), most incorrect answers were based on adding fractions rather than multiplying them. Many of those candidates who knew to multiply the fractions were let down by an inability to do so.

1.2.15 Question 15

Most candidates found this question challenging. In part (a), only the best candidates were able to calculate the moving averages and plot them all on the graph. Very few knew that they should be plotted at the midpoint of the values. In part (b), candidates seemed to be unsure about how the moving averages display the trend of the gas bills many referred to "ups" and "downs", or "fluctuations", in the gas bills. In part (c), only a small minority knew about seasonal variations, but some gained a mark for making a sensible suggestion as to how they might be caused.

2 PRINCIPAL EXAMINERS REPORT – PAPER 1H (FOUNDATION)

2.1 GENERAL COMMENTS

The paper seemed to be well received by candidates. Virtually all the candidates managed to get at least 25 marks. Almost all the candidates completed the paper. A few had some spare time judging from the writing on the back page although this was rarely the case with the able candidates. All the questions were attempted by most of the candidates.

2.2 REPORT ON INDIVIDUAL QUESTIONS

2.2.1 Question 1

Part (a) This was not a problem for the majority of candidates. Part (b), half the candidates were able to come up with figures for stratifying (9 black and 1 brown), but omitted to say how these would be chosen. Half the candidates mentioned random sampling. It was quite rare to see both. Most candidates scored 1 mark out of two.

2.2.2 Question 2

The majority of candidates did well on this question, scoring at least one mark and, quite commonly, 2.

2.2.3 Question 3

Part a)(i), it was rare for this part of the question to be done incorrectly,(ii), few candidates recognized this as a conditional probability question. The most common answer was 3/30. Part (b), most candidates realized that the data did not support the idea. Many candidates used only a vague reference to the data or referred to the first column, thinking that those that had cake were representative of all 30 people.

2.2.4 Question 4

Part (a), this was mostly done well. Some candidates incorrectly superimposed the bars while others, equally incorrectly, had them side by side. Part (b), the vast majority of candidates had little difficulty with this question. A few had problems with the arithmetic even though they had the right method. Others included females as well as males, having misread the question.

2.2.5 Question 5

Part (a), candidates did not find this an easy question. Many felt that the *x* axis was confusing, candidates should be familiar with this as a common format for non calendar years (e.g. academic, as here, or financial). A large number of candidates referred to the *y* axis showing % rather than a number, a feature that is not misleading but which was the crux of part (b). Others thought that

this axis going up in twos was misleading. Less than half the candidates picked up the *y* axis starting at 2.
Part (b), many candidates got this correct especially when they had referred to it in part (a).

2.2.6 Question 6

Part (a), most candidates were able to deal with index numbers. This was an all or nothing question in terms of marks. If the candidates knew the method they got the right answers.
Part (b), few candidates were able to tackle weighted averages. Again this was all or nothing in terms of marks.
Part (c), although a simple follow through from part (b) was expected, the majority of candidates did not see the simple connection with the answer to part (b).

2.2.7 Question 7

Part (a), this was quite well answered. Occasionally candidates were rather vague referring to 'most' games having been won. Some candidates thought that 00-49 contains only 49 numbers, thus demonstrating a lack of familiarity with simulation. Part (b), most candidates did this correctly.
Part (c) (i), most of the candidates that got part (b) correct were able to do this part of the question. Some gave vague answers such as 'badly' or 'they don't', and as a result lost a mark,(ii), many candidates were unable to express themselves clearly making vague references to 'more random numbers' or 'better numbers' and giving the impression that they meant the range of numbers allocated, rather than doing more simulations. Others thought that the numbers allocated to winning losing and drawing should be reallocated so that the probability of winning was increased.

2.2.8 Question 8

Part (a), a lot of candidates only identified the word 'census' with the 'National Census' taken every ten years. This led them to give incorrect reasons why a census should not be used. Quite a few candidates correctly identified cost and/or size as being reasons.
Part (b), quite a number of candidates gave answers which lacked sufficient specification, referring just to 'residents' or 'the city'. Other candidates considered that the population should consist of drivers only.
Part (c), most candidates had the right idea but some found difficulty in expressing it. A minority thought that closed questions would avoid bias. Some candidates thought that they result in Yes/No answers.
Part (d)(i), most candidates got this part right, although few gave the most succinct answer that it was a 'leading' question. In (ii), most candidates picked up at least 1 mark here. Some candidates gave a question that was equally leading – e.g. using 'agree' without using 'disagree' or 'good idea' without 'bad idea'. Some centres

appeared to have not got across the use of tick boxes for closed questions.

2.2.9 Question 9

Part (a), this was generally done well. Some candidates thought that $(-4)^2 = -16$. There was also a tendency to use $n(n - 1)$ rather than $n(n^2 - 1)$. Part (b)(i), this was generally answered well by many candidates including those that got part (a) wrong. In (ii), the meaning of rank correlation was not well understood. Many candidates went back to the original data and referred to the teams having similar points or abilities.

2.2.10 Question 10

Part (a),(b),(c),(d) and (e) - All these sections were done well by the majority of candidates.

1.2.11 Question 11

Part (a), a number of candidates had all the points to one side of the line or drew a line that went through the middle point with the points to the left above the line and those to the right below. Given that the points were nearly in a straight line this was rather surprising. Part (b), it was exceptional for marks to be lost in this part. Part (c), most candidates were able to say something sensible here. Part.(d), this was not done well. A large number of candidates tried to use the last value of 105, either adding (or subtracting) the seasonal value from this, or adding 5 to it on the grounds that it had gone up 5 the previous year.

Part (e)(i), even those that got part (d) right had difficulty in coming up with the 'text book' answers of continuing trend and seasonal effect. In (ii), the answer was dependent on getting part (i) correct.

2.2.12 Question 12

Part (a), although many candidates knew how to find the quartiles some were not able to actually complete the process.

Part (b), a number of candidates answered 55 without showing any working. The answer could easily have come from guesswork. Working must be shown. A number of candidates correctly found 1.5 × IQR but then worked from the median rather than the quartiles.

Part (c), few candidates seemed to be aware of how to draw box plots with outliers. Part (d), there were some good answers although it was obvious a number of candidates did not understand the question. Some thought that 'symmetric' was the name of a distribution, others just gave the reason without naming a distribution. Part (e), this was done well by all candidates.

2.2.13 Question 13

Part (a) (b), few candidates seemed to have grasped the basic idea of quality control, and that the Normal distribution was being used. Quite a few quoted 99.8% and 95% without being able to relate these to the questions.

Part (c), many candidates did not seem to have grasped that the idea of quality control was to control the process as it is being carried out and that it is a check on the output of the machine so that if the population mean changes it can be stopped and reset.

Part (d), many candidates thought that 'Take another sample' was the answer. There were quite a few stopping the machine but few resetting.

Part (e), this was answered well, most candidates realising that Bob would be losing money.

Part (f), few candidates realised that the control chart for weight was set up using the standard deviation so that this (or the range) would also need to be checked.

2.2.14 Question 14

Part (a), most candidates did this well.

Part (b), few candidates realized that they could do this part by subtracting from 360^o, but it was still done correctly in most cases.

Part (c), the majority of candidates did this well and most included suitable labelling.

Part (d), some candidates thought that because the rank of the sectors were the same that this was a sufficient reason, most however realized that it was the proportions that they had to look at.

2.2.15 Question 15

Part (a), this was done well by most candidates.

Part (b), a number of candidates obviously did not know how to use the tree diagram in order to answer this part of the question. This was most disappointing for higher tier candidates.

Part (c), this was not well done. Again many candidates did not know how to use the tree diagram. Some used the conditional probability formula but found P(late $|$ given bus) rather than P(bus $|$ late).

Part (d), this should have been a simple follow through mark but many candidates did not realize that P(did not catch bus) = 1 – P(caught bus).

2.2.15 Question 16

Part (a), this was done well by the majority of candidates although there was a tendency to miss out half the pairings.

Part (b), although a number of candidates gave the correct answer it was surprising how many did not, considering that this is likely to have been done in KS3 Maths.

Part (c), this was not done well. Some candidates gave non integer answers.

Part (d), this was not done well. Often all the terms of the expansion were used to calculate the answer. A few candidates used 7 and 2 or 0.7 and 0.2 instead of 7/9 and 2/9 when calculating this answer. It should be obvious to candidates that probabilities > 1 are wrong.

2.2.16 Question 17

Part (a), higher tier candidates should be able to do basic histograms and the mean of grouped data. This was often not the case and many candidates did not use frequency density in this part of the question. A few candidates had difficulty deciding on a suitable scale.

Part (b), a lot of the errors in this part of the question were due to difficulties in finding the mid-points – the mid point of the interval 6 – 12 was frequently given as 9.5. It is fairly simple to add the two numbers together and divide by two.

Part (c), there were some problems here although these were less of a surprise. Candidates should be encouraged to use the 'calculation formula' rather than the 'definition one'. A common error was the use of 2872 for the Σfx. Candidates often disregarded the instruction to give the answer to one decimal place and lost 1 mark.

3 PRINCIPAL MODERATORS REPORT – 1389/02 COURSEWORK

3.1 GENERAL POINTS

Most centres sent their marks and the requested coursework to their moderator by the date on which it was due. A few centres caused problems by failing to meet this deadline.

3.1.1 Moderators were helped by centres who attached a completed task sheet to the front of each candidates work. submitted work which was securely fastened using treasury tags provided annotation or comments as to why and for which aspects of the work they had awarded marks.

3.1.2 Some centres submitted work which was not adequately fastened to the task sheet or didn't send the task sheet at all. Some task sheets lacked candidate numbers or had incorrect candidate numbers. These administrative failings seriously hindered the moderation of the coursework.

3.2 Coursework Tasks

There were a few centres that chose their own project; collecting their own primary data, often as a group effort and then comparing with some secondary data from an internet site. Having more ownership of the project encouraged Candidates to explore their own ideas and enjoy their work rather than having an attitude of 'this is something we have to do let's do it with the least possible effort'. The few schools who allowed their candidates to follow their own interests producing individual projects tended to achieve excellent results.

This year 'Cars' actually seems to be the easiest for higher candidates to score highly on. The limited database means that whilst sampling techniques can still be demonstrated, the amount of time spent on this stage is not disproportionate – which is definitely the case with Newspapers. The structure of the assessment guidelines into strands and substrands enabled most candidates to produce work which met the criteria.

3.2.1 Strand 1a Planning

Most candidates had some sort of plan and usually managed to write down a hypothesis but detail was missing even in the better work. Candidates made predictions but few set out their strategy for investigation in any detail. A discussion of why they chose to use the different techniques to investigate their data is essential. Centres must be aware that for a mark of 4 or above the task must be sufficiently complex to give candidate opportunities to make comparisons and to explore interrelationships between the variables investigated. Many centres limited the planning marks available to their better candidates by giving too much help in planning. A list of instructions telling candidates what to investigate and which techniques to use is undue help for candidates above foundation level.

3.2.2 Strand 1b Collecting

Choosing the best sampling method is essential in this strand. The methods of sampling were rarely fully described. Many candidates had been told that they had to include information on lots of sampling techniques. It was not uncommon to see pages copied from books but no information on what the candidate actually did. A number of centres awarded 4 marks simply because candidates had used stratified sampling. Those who did describe the sampling often only went half way. Stratified sampling was considered and numbers worked out but no method was given for selecting from within the strata. To just say 'random' was not enough. Candidates should describe exactly how the random button etc was used. In a number of cases the candidate took proportional samples e.g. 10% from different year groups in Mayfield High and then compared between them! This was then marked as correct stratified sampling!!!

More able candidates need to plan for anomalies and outliers including formal methods of identifying them. A more structured approach to these should be encouraged.

3.2.3 Strand 2a Analysis

This strand was done well by the majority of candidates. The less able need to understand that titles/axis labels are needed. The more able need to explain why they have chosen to use certain techniques and diagrams. Comparative diagrams were often not done on the same page by middle to lower ability candidates. Some of the able candidates produced some very nice overlays that worked very well.

The main graphs used were scatter graphs and box plots. The scales of these were not always set up for comparison. Some candidates attempted to analyse the box plots in terms of comparing the key statistics, leading on to comments on skewness. A few centres made the mistake of giving credit to candidates who drew graphs of person number against height. The majority of candidates used Excel or other computer packages to draw graphs and calculate any statistics they wanted. They must explain the meaning of anything they find if they are to gain credit and to show that they can understand how to use sensible scales.

3.2.4 Strand 2b Calculations

It was often very difficult to find any evidence of calculations in the weaker work. Some candidates used scatter graphs alone and no calculations at all. Able candidates usually, managed to calculate the medians and quartiles and often Spearmans Correlation Coefficient. Centres should encourage candidates to plan hypotheses so that a variety of appropriate graphs and calculations can be used in the project. Many candidates calculated Spearman's Correlation Coefficient but didn't use it. Even at the upper end of the ability range very few candidates scored 9-10 marks. Standard Deviation unless used for a specific reason is not worth more marks than the interquartile range.

All candidates should realise that a computer will work out all sorts of statistics, but these can't be given credit unless they are used and interpreted meaningfully. Using a statistical package to generate a range of statistical values which are stated but not used constitutes redundancy and detracts from higher level marks.

3.2.5 Strand 3 Interpretation

Almost all candidates scored at least 4 marks as they could all comment correctly on the correlation and what that meant in their project. The more able tried to pull everything together in a conclusion. Some candidates produced some creditable work using the standard deviation and the normal distribution. It was good to see some candidate deliberately not working out Spearmans Correlation Coefficient if there was obviously no correlation. One or two candidates also discussed causality rather well when dealing with correlation. Generally the maximum mark achieved was 9 and that was rare. Candidates need to be encouraged to think more deeply about what their results mean in the real world. They should consider what might be happening and why. This should undoubtedly be a hallmark of A grade candidates.

3.3 Concluding Remarks

Many candidates are still producing far too much work. The essence of the project is to choose the correct techniques and apply them. Additional pages of bar charts and pie charts which add nothing to the project should be discouraged as they detract from the elegance of the project. In conclusion. It is obvious that some of the best work would not be out of place at A level and that this course is improving the understanding of statistics in this age group.

4 STATISTICS

4.1 MARK RANGES AND AWARD OF GRADES

Unit/Component	Maximum Mark (Raw)	Mean Mark	Standard Deviation	% Contribution to Award
1389 / 1F	80	47.8	12.9	75
1389 / 1H	100	60.9	16.7	75
1389 / 02	40	21.5	6.2	25

4.2 GRADE BOUNDARIES

The table below gives the minimum raw marks required for each component grade

	Max	**A***	**A**	**B**	**C**	**D**	**E**	**F**	**G**
1F	80				**52**	42	33	**24**	15
1H	100	81	**66**	51	**36**	**25**			
02	40	30	**26**	22	**18**	**15**	13	**11**	9

4.3 OVERALL GRADE BOUNDARIES

The table below gives the minimum subject marks required for each overall grade.

	A*	**A**	**B**	**C**	**D**	**E**	**F**	**G**
Foundatin				59	48	37	26	15
Higher	79	65	51	38	28			

Further copies of this publication are available from
Edexcel Publications, Adamsway, Mansfield, Notts, NG18 4FN

Telephone 01623 467467
Fax 01623 450481

Email publications@linneydirect.com

Order Code UG017075 Summer 2005

For more information on Edexcel qualifications, please visit www.edexcel.org.uk/qualifications
Alternatively, you can contact Customer Services at www.edexcel.org.uk/ask or on 0870 240 9800

Edexcel Limited. Registered in England and Wales no.4496750
Registered Office: One90 High Holborn, London, WC1V 7BH