Analysis of Student Performance in Statistics

Narelle F. Smith and Leigh N. Wood
School of Mathematical Sciences
University of Technology, Sydney

Roslyn K. Gillies
Student Learning Centre
University of Technology, Sydney

Gillian Perrett
Faculty of Education
University of Sydney

Abstract

This paper reports on an investigation of students' difficulties in first-year statistics examinations at university. Our hypothesis was that difficulty with language was an important factor in student performance in statistics examinations. Our data consisted of examination questions ranked in order of difficulty based on student performance, lecturers' perceptions of difficulty and a measure of the linguistic complexity of the questions.

The examination paper that was analysed was a typical short answer paper for students studying statistics in their first year at university. The paper was not designed specially for analysis but rather was the normal end of year paper for a business statistics subject with an enrolment of 600 students. Three statistics lecturers not involved in teaching the subject ranked questions according to their perception of the level of difficulty and these were compared with the performance of 186 students. The examination paper was also analysed for linguistic complexity as measured by lexical density and this was compared with students' responses.

The results were surprising. There was no correlation between student performance and the linguistic complexity of the questions as measured by formal measures of lexical density. The lecturers' rankings were consistent and correlated highly with student responses in most cases . Certain topic areas appeared to cause more difficulties than others and further research will concentrate on these topics.

Introduction

This paper reports on an investigation of students' difficulties in first-year statistics examinations at university. Our hypothesis was that difficulty with language was an important factor in student performance in statistics examinations. Answering statistics questions involves the skill of encoding and decoding from the English language to the language of concepts of statistics. Key words and phrases are used to signify a particular statistical situation and that a particular technique is required to solve the problem. As well, statistical questions are quite rightly generally set in real world situations. Much background information is given to set the scene; and consequently a further skill is required of sifting through the information given and extracting that which is relevant.

Staff in university mathematics learning centres have assisted students with statistics and after many interviews identified language as one of the key areas of difficulty for students of statistics. In many cases, the students have been well able to handle the mathematics involved but have been unable to interpret the English language question in terms of their statistical knowledge. Examples abound. Students may be able to construct or find a 95% confidence interval but be unable to find the limits between which 95% of values are expected to lie. Students may understand exactly what "less than or equal to" means, but be unable to understand "not more than". They may be able to perform the mechanics of a hypothesis test, but be unable to perceive that an advertising campaign is a success if we can show that sales have increased.

Ideally, the goals of introductory statistics courses should involve understanding, acquisition of higher order thinking skills and appreciation of the relevance and utility of statistics (see Williams, 1993). The attainment of such goals must involve a recognition that language skills play an important role in statistics.

As a first step in investigating the importance of language in statistics, questions on first year statistics papers were analysed for linguistic complexity. The results were compared with the actual performance of students on these questions and with a ranking for each question based on lecturers' perceptions of difficulty.

Method

Subjects
Students in their first year of a Business degree study a statistics subject. This project analysed 186 examination papers which were all the papers of students from one campus of the University. There were very few students who spoke languages other than English as their main language. A majority of students were straight from school.

The questions were ranked on perceived levels of difficulty by three statistics lecturers at two universities who knew the subject area but did not teach the subject. They were asked to rank the examination questions using a 5 level difficulty scale, from 1 - easiest to 5 - hardest. Median rankings were used for each question.

The examination
The examination was the normal end of semester examination for the subject and was an 'open-book' examination. The exam was broken into 40 questions mostly requiring short answers. Marks for each section were analysed for numbers of students who received full marks and numbers who received no marks for each section.

Analysis of linguistic complexity
There is no single measure of linguistic complexity. Simplistically it may be assumed that for any student who has difficulty with language the basic amount of language that has to be processed in order to answer a question may be the cause of difficulty. Other measures include a calculation of the lexical density of the questions. Measures of lexical density are commonly used as easy ways of

distinguishing spoken styles of language from written styles and measuring the difficulty of a written text. Such a measure was used by Davies (1991) in his review of the NSW Basic Skills Test of numeracy which discusses issues relating to the linguistic difficulty of numeracy questions for NESB primary children. The measures examined here were:

- number of words in the question;
- number of clauses needed to be read (using definition from Halliday, 1985:67);
- Lexical density 1. Lexical density can be expressed as the ratio of lexical words to total words in a stretch of text (Halliday 1985:64). For example: "*Neatly sketch* and *label* a *stem* and *leaf display* of the *data*" contains 7 lexical words (italics) and 5 grammatical words. The lexical density would be 7/12 or 0.58.
- Lexical density 2. Another measure of lexical density is the average number of lexical words per clause in any stretch of text. It is the approach preferred by Halliday (1985), Davies (1991) and Halliday and Martin (1993). Halliday (1985:66) explains

  > [Density] has to do ... with how closely packed the information is.... It is this packaging into larger grammatical structures that really determines the information density of a passage.

  Halliday (1985) describes texts with a high proportion of lexical words as lexically dense and it is often assumed (Davies 1991, Halliday & Martin 1993:69) that texts which are extremely dense lexically are difficult to understand, especially for non-English speaking background readers. Halliday (1985:77ff) argues that texts low in lexical words are correspondingly more grammatically intricate.

Results

Performance of students

The examination paper was structured so that each topic in the course was examined and the order of the questions roughly represented the order in which the topics had been taught. Overwhelmingly, the questions handled best by students were the first three which were straightforward applications of exploratory data techniques - stem and leaf display, calculation of median and quartiles and sketching a boxplot. In fact 184 out of 186 students (99%) correctly sketched and labelled the stem and leaf display.

Apart from the good performance on the first introductory topic, there was no pattern of decreasing performance as the paper progressed. The students did not appear to be more proficient with topics introduced early in the course nor did their performance deteriorate through the duration of the examination.

The questions answered most successfully were interspersed throughout the paper and fell into two categories: straightforward calculations using a formula or questions relating strongly to topics covered in school mathematics such as plotting data from a chart or interpreting the meaning of the slope of a regression line. The last of these was of particular interest. As a question involving interpretation skills, it was ranked as a 4 on the five point scale of difficulty by the lecturers. However, 153 of the 186 students (82%) answered this question correctly. This suggests that the students have a real understanding of the concept of slope and gradient from their school mathematics.

The most poorly handled questions were conditional probability, hypothesis tests, exponentially smoothed forecasts and interpretation questions on control charts - all concepts/techniques new to students and not based on school mathematics to any great degree. The question on conditional probability - a relatively straightforward application of Bayes Theorem dealing with posterior probabilities - was only answered correctly by 11 students (6%) and was either not attempted or scored no marks at all by 139 of the 186 students (75%). No student scored full marks for the matched pairs hypothesis testing and 98 students (53%) scored no marks at all or did not attempt this question. A copy of the questions on conditional probability and matched pairs hypothesis testing is in Appendix B.

Comparison of performance with difficulty ranking

There was broad agreement between the lecturers on level of difficulty of each question and also between median difficulty rank on each question and the actual student performance. An analysis of variance, explaining variation in the number of students answering the question correctly and the median difficulty rank for the corresponding question was significant ($p = 0.000$). In general, the lecturers considered that routine questions (using a standard technique) were less difficult than questions involving interpretation skills and this was reflected in the students' results. For example, 41% of students gained full marks on a Normal Distribution question that had been graded as '2' and 24% for a question on the same topic graded '3' by lecturers. Overall, the lecturers appeared to be "in tune" with the students.

There were some discrepancies between the actual performance and the lecturers' rankings: notably, a question involving interpretation of the slope of a regression line (previously mentioned) and the questions on hypothesis testing. Performance on the regression question was markedly better than the difficulty ranking would suggest. The hypothesis testing questions were poorly performed but were not uniformly ranked by the lecturers as very difficult.

Comparison of performance with linguistic complexity

Two measures of lexical density were calculated and compared with performance as measured by the numbers of students achieving full marks and the numbers receiving no marks. The correlations between numbers of students scoring no marks and the two measures of lexical density were 0.139 and 0.094 respectively. These correlations are very low and indicate that these measures of linguistic complexity did not explain variations in student performance on different questions.

Because of the relatively homogeneous nature of this group, we were not able to look at the performance of particular groups of students, such as non-English speaking background.

We need to be a little wary about using lexical density in these analyses even though it has been used previously to analyse mathematics examinations (Davies, 1991). The generally held hypothesis that high lexical density is an important cause of difficulty may not apply in this situation. A lexically dense phrase such as *finite population correction factor* is taught as a technical term to students of statistics and may simply indicate that a particular formula should be used. Further, words of crucial significance in a statistics question such as "less than" are regarded as grammatical and not lexical in a pure linguistic analysis. It appears that lexical density may not be the appropriate tool to test linguistic complexity in statistics questions.

Discussion and Conclusions

For our group of students, performance was not explained by the timing of the topic in the course or by the position of the question in the examination. Lexical density, as a formal measure of linguistic complexity, did not explain variations in student performance on different questions. As expected by the lecturers, students performed best on questions that required mechanical calculations (see Williams, 1993).

Specific topic areas appeared to cause the most problems, and our personal experience with individual students seeking remedial assistance suggests that difficulty with the coding and encoding from the English language to the mathematics plays an important role. One conclusion from this work may be that this encoding/coding skill is independent of the linguistic difficulty of the questions as measured by lexical density. As the next step, our intention is to study the individual examination scrips with particular reference to the questions on conditional probability and hypothesis testing, in an attempt to determine what aspects of these topics cause students the most problems.

References

Davies, A. 1991. Performance of children from non-English speaking background on the NSW Basic Skills Tests of Numeracy: Issues of test bias and language proficiency. *Language, Culture and Curriculum, 4*,2,149-161

Halliday, M.A.K. 1985. *Spoken and Written Language.* Geelong, Deakin University Press

Halliday, M.A.K. & Martin, J.R. 1991. *Writing Science: Literacy and Discursive Power.* London, The Falmer Press

Williams, A, 1993. Is a Pass Good Enough in Tertiary Statistics?, *Proceedings of MERGA-16*, pp 587-591.

Appendix A

Question Number	Lexical Density 1	Lexical Density 2	Difficulty Ranking	No. of Students with full marks
1	0.52	5.00	1	184
2	0.48	4.67	3	131
3	0.49	5.00	2	164
4	0.53	4.62	2	67
5	0.40	3.57	4	11
6	0.38	2.40	2	61
7	0.47	4.00	2	77
8	0.43	4.33	3	45
9	0.34	3.33	1	165
10	0.36	3.57	3	57
11	0.32	3.17	3	116
12	0.34	3.14	3	110
13	0.47	3.40	3	123
14	0.35	3.50	3	81
15	0.52	4.80	4	82
16	0.52	4.06	4	27
17	0.46	4.18	3	57
18	0.49	5.44	4	112
19	0.49	5.17	4	0
20	0.51	6.14	3	92
21	0.53	5.00	2	126
22	0.40	4.00	4	153
23	0.51	4.80	2	119
24	0.52	5.17	2	80
25	0.52	5.18	3	100
26	0.53	4.69	4	81
27	0.47	4.83	5	45
28	0.56	6.33	5	53
29	0.62	4.00	3	98
30	0.62	4.00	3	95
31	0.53	4.00	2	146
32	0.55	4.50	4	99
33	0.56	4.14	4	57
34	0.42	4.00	4	43
35	0.50	9.00	4	117
36	0.47	5.33	5	51
37	0.41	2.25	4	87
38	0.53	5.00	5	25
39	0.53	5.25	5	12
40	0.51	4.90	5	11

Appendix B

Question 5

a) The producers feel that the new movie has a 30% chance of being a success. A certain film critic, who has liked 70% of all successful films and disliked 80% of all unsuccessful films that she has reviewed, is going to review the film. Find the probability that the movie will be a success given:

(i) The critic likes it.

(ii) The critic dislikes it.

Questions 18 - 20

In order to measure the effect of a storewide sales campaign on non-sale items, the research director of a national supermarket chain took a random sample of 11 pairs of stores that were matched according to average weekly sales volume. One store of each pair (the experimental group) was exposed to the sales campaign, and the other member of the pair (the control group) was not. The following data indicate the results over a weekly period.

Sales ($000) of Non-sale Items

Store	With Sales Campaign	Without Sales Campaign
1	67.2	65.3
2	59.4	54.7
3	80.1	81.3
4	47.6	39.8
5	97.8	92.5
6	57.3	52.4
7	75.2	69.9
8	94.7	89.0
9	64.3	58.4
10	31.7	33.0
11	54.0	53.6

(i) Is a one or a two tailed hypothesis test more appropriate here? Justify your answer briefly and use your selected approach in (ii) below.

(ii) At the 0.05 level of significance, can the research director conclude that there is evidence that the sales campaign has increased the average sales of non sale items?

(iii) What assumption is necessary to perform this test?