Understanding of Association and Regression by First Year Economics Students from Two Different Countries as Revealed in Responses to the Same Examination Questions

John Truran
University of Adelaide

This paper addresses the broad issue of relating research findings with pedagogical practices by analysing the responses to questions set in an undergraduate statistics examination using Eisner's connoisseurship and criticism approach, supported by general pedagogical and psychological principles. Comparisons are made between responses to the same course given in two different countries to assess similarities, differences, and weaknesses in order to indicate possible ways in which future courses might be modified to improve student learning.

The Purpose of this Investigation

There is currently an agreement between the University of Adelaide, Australia, and the Sepang Institute of Technology, Malaysia, which permits Malaysian students to prepare and sit for examinations at the University of Adelaide by attending lectures and tutorials at Sepang. One such course currently available is "Business Data Analysis" (BDA) which is a compulsory first-level course in basic statistics undertaken mainly by students studying for a Bachelor of Economics or Bachelor of Commerce degree. Currently it is possible for Malaysian students to take the first two years of their degree at Sepang, after which they come to Adelaide to complete their studies and receive an Adelaide degree.

As far as possible, the course offered in Sepang matches that in Adelaide. For the semester considered in this paper the Adelaide lecturer in charge of the whole course kept a firm hand on all aspects of the course, her notes and instructions to both students and staff were detailed and precise. The Malaysian lecturer received all of these notes as a basis for his lectures, his students had the same amount of contact time, and did the same exercises, assignments and examinations. Tutorial exercises, which counted for the final assessment, were marked in Malaysia, but special assignments and examinations were marked by the teaching and tutorial staff in Australia.

In second semester, 1996, a total of 272 Adelaide students and 88 Malaysian students sat the final examination. Observation of the author's three tutorial groups suggests that the Adelaide class is fairly diverse, with representatives from most major cultural sub-groups in South Australian society and with about 10% of its members being student visitors to Australia, mainly from a number of different Asian countries. Little is known about the Sepang group, but all the students have Chinese names, so it may be presumed that they are predominantly Malaysians of Chinese origin.

Such an arrangement provides about an excellent opportunity to compare the approaches of two quite different groups to an almost identical course in order to see if there are significant differences in understanding or approach which need to be considered in designing the presentation or examination of the subject. In order to match findings with other work it was decided to investigate similarities and differences in students' responses to those questions in the final examination which were concerned with the ideas of association and linear regression.

Methodological Principles

Analysis of examination scripts is rarely reported by researchers. It has limitations, including an inability to follow up answers in more detail, and, as Lipson (1994) has clearly shown, a tendency to assess product rather than process learning. On the other hand, students have a strong motivation to present their best work, so it may reasonably be claimed that the answers provided represent a good estimate of what they consider to be relevant to the questions as presented. Furthermore, questions tend to be precisely

formulated and clearly stated, students know that the marking scheme is additive, so there is no penalty for incorrect statements, and usually they quickly discover that they have adequate time to answer the questions thoroughly. For the researcher the method is time-efficient, and is relevant to classroom practice because it has the potential to improve both the teaching and examining processes.

On the other hand, it is not possible for a researcher to obtain further information to clarify what has been written, or to assess by alternative approaches whether similar responses from different students really do represent similar understandings. There may well be inconsistencies within an answer which may represent either a slip or an important misunderstanding. And it is often easy for phrases learned merely by rote to give the appearance of sound understanding. This is of special importance in a subject like BDA where much of the work comprises routine processes and routine interpretations.
But the principles of textual analysis have been developed to cope with just these situations. They have, however, rarely been applied in mathematics or to examination scripts. Clearly such analysis needs to be done cautiously and with a clear sense of the provisional nature of any conclusions which are drawn. The approach employed here follows Eisner's (1985, ch. 6) proposal for using *connoisseurship* and *criticism* in order to make helpful evaluations. Connoisseurship he sees as an appreciative and very personal art; criticism as a way of describing what is present so that others may enter into the work, and make their own judgements. In this paper a mixture of systematic elementary data analysis and broad mathematics education principles are used to examine the data in order to determine which aspects of the teaching of one main topic have been successful or unsuccessful, and to provide a basis for subsequent modification of the course. In other words, both qualitative and quantitative analysis are seen as relevant.

Such an approach fits within the evolutionary model of curriculum change proposed by Popkewitz (1988, pp. 242 - 244) which has limitations, but can be very efficient for implementing minor changes quickly, and for strengthening good practice. It is of special value when limited resources particularly encourage the use of time-efficient forms of assessment like written examinations and is able to estimate quickly the discriminatory power of the questions and to suggest appropriate improvements.

The Course

The course concerned, "Business Data Analysis", is a compulsory semester-length course, taught by one lecturer for 2 hours per week. Each student should attend a weekly one hour tutorial class usually with about 10 participants, taken by a number of tutors, of whom this author has been one. In practise about one-third of the Adelaide students do not attend tutorials. The figures for Sepang are not known. Tutors mark assignment questions, and discuss work which the students are expected to prepare for each tutorial. Basic principles and expectations are reiterated many times in lectures and tutorials. Students are encouraged, both verbally and by the mark schemes employed, to memorise appropriate language for reporting statistical interpretations.

Most of the teaching staff concerned with the course do not have any formal training in education, and their pedagogical principles tend to be traditional. This does not imply that they do not do their job well, but rather that they see their job as the transmission and retention of information and skills. The constructivist position that there is no mathematics "out there", but only in the mind of the learner, would not be understood by most of them. So the analysis in this paper cannot assume technical understanding of educational theories of any form. This makes the connoisseurship approach particularly appropriate, because if the analysis is to be of any value it must convince intelligent people who are "outsiders" to modern educational thinking, while remaining consistent with sound principles of mathematics education.

One aspect of the conflict between constructivism and traditional approaches is of special importance here. Within the constructivist position there is no such thing as a "misconception", but within the transmissionist position misconceptions are one of the most important features that a teacher needs to identify and try to eliminate. The position taken in this paper is that misconceptions do exist and it is useful to identify the existence and prevalence of significant understandings and misunderstandings by students.

Methodology

Many misconceptions are not widely held, but are still important, so aberrant responses are of special interest. Therefore it was decided to examine a relatively large number of scripts. In Adelaide 272 students (90% of the enrolled students) sat the examination and half of these scripts were analysed, selected in chunks from the full alphabetical ordering more with regard to physical convenience than to any other criterion. This was considered adequate for the precision required by the research. All 88 Malaysian scripts were analysed in the same way as the Adelaide ones.

All the data were entered onto an electronic spreadsheet. For each part of each question the different responses from students were listed on the horizontal axis. The electronic spreadsheet approach allows the headings to be built up as the scripts are read. Clearly, the greater the number of scripts the greater the number of different responses there is likely to be. Even so, building up the responses is not easy, because it may not be obvious at the beginning that certain aspects of the responses are of special importance. It will be seen below that on a couple of occasions inadequate data were recorded. Given the very limited amount of previous research done into these topics, this is not surprising. On the vertical axis a code for each student was entered. Each cell in the table was filled in with either "1" or "0" or "blank", indicating a correct, incorrect, or absent response, respectively. This approach is consistent with the reality that new classifications are being added to the table as the data set is being built up.

The relevant parts of the one question examined are presented here. While no choice of questions was permitted, it was still clear that this question was popular, because it was frequently attempted early in the examination, and few students seemed to be under time pressure when attempting it.

Question 4 presented a summary of information illustrating a relationship between the age of a motor cycle and its resale value, and asked students to calculate measures of association and the simple linear regression. Formulae were provided; the calculations were usually done correctly. The following parts are analysed here (sections in italics were not examined but are included here to clarify the context of the parts which were considered):

(a) *Calculate and* interpret both the correlation coefficient and the coefficient of determination.

(c) Interpret both the intercept and slope of [the regression equation you have calculated].

(d) Predict the selling price of a motor cycle that is
(i) 3 years old
(ii) 6 months old.

(e) What reservations, if any, do you have about prediction (i) and prediction (ii) in part (d).

Two major techniques of analysis were employed. The first involved simple analysis of the relative frequency of each response encountered. This was useful for identifying the popular responses, and weak areas in the group's responses as a whole. This, together with experience gained while teaching the course, was helpful in indicating where responses tended to have been rote learned, and in identifying subtle differences in students' understanding of similar ideas. The second major technique used electronic copies to re-order the columns so that similar ideas in the same question or in different questions were adjacent. The data could then be sorted electronically into numerical order by row entries (1, 0, blank) to identify, for example, which students had a "1" entry for, say, all the similar ideas, or for various subsets of the ideas.

Previous Research into Understanding of Association and Linear Regression

The concepts of statistical association and linear regression are fundamental to statistical thought. Calculating regressions is straight-forward, but my experience is that the underlying ideas sometimes prove difficult, a view supported by Franklin (1988) and

Tamura (1994). A systematic search of *Teaching Statistics* and *Journal of Statistics Education* found a number of articles presenting interesting ways of teaching both regression and correlation (Goode & Gold, 1987; Laviolette, 1994; Wilkie, 1980). These suggested, either implicitly or explicitly, that using such ways would improve students' understanding of the concepts, but they did not specifically identify students' difficulties or misconceptions. The emphasis was on *teaching,* rather than on *learning,* statistics. Indeed, there has been little done in this field (Batanero et al., 1994, p. 538; Shaughnessy, 1992; J. & K. Truran, 1996). It has been observed that students often have difficulties in interpreting a single measure of correlation, especially in elementary courses (Hawkins et al., 1992, pp. 51 - 56). Students' understanding of correlation as represented in contingency tables has been analysed in detail (Batanero et al., 1996) and tertiary students' understanding of the significance of r has been analysed briefly and somewhat imprecisely (Truran, 1995) using a method similar to that described here.

Results

The purpose of this survey is partly to inform the teaching of the BDA course in subsequent years, so the results presented here consider the most striking differences between the two groups and/or those situations which most clearly indicate important misconceptions which need to be addressed in teaching the topic. Where percentages are given as ordered pairs, it is the Adelaide figures which are given first.

Question 4 (a) - Interpretation of r

The correct value of r was -0·57. Students were expected to indicate that this indicated a moderate, inverse, linear relationship between the variables. Only correct responses for r were analysed further (84%, 89%). Of these, almost all students appreciated the significance of the negative sign (93%, 97%). A significant minority of both groups failed to observe that the relationship was linear (27%, 24%) In both cases the differences between the groups were small.

The importance of stating that the relationship was both inverse and linear was emphasised throughout the course, and in general both were treated together. The striking difference between the students' willingness to state both of these needs to be considered. While we have no evidence about how these students thought, it is possible to use principles from mathematics education to explain the discrepancy. Skemp (1971, ch. 2) has argued that in the development of a concept it is important to be clear about its boundaries. Since the students have met examples of both positive, negative and zero association between variables, they have been able to see the importance of this classification. But the students have met only linear associations in any detail, and have not encountered the importance of deciding whether such a model is appropriate at the beginning.
The difference in success rates for two items treated together indicate that emphasis in class and rote learning do not necessarily produce correct responses in examinations, and this effect seems to be similar across both groups.

Assessing the strength of the relationship requires and expression of an opinion, and there were definite differences in the groups' perceptions as may be seen in Table 1. (Responses like "moderate to weak" were classified by the first term stated.)

Table 1
Verbal Meaning for r = -0Æ57

Response	Adelaide	Sepang
Strong	0%	8%
Fairly strong	4%	6%
Moderate	61%	47%
Weak	25%	22%
No term used	10%	17%

In so far as any language has meaning across cultures, the Sepang students tend to see the relationship as stronger than the Australians do. This result shows that vernacular terms are unreliable instruments for describing the strength of a relationship. The cross-cultural differences and the fact that no student indicated considering the number in the

sample means that there is a need to provide students with a less individualist way of making judgements about correlation. A significance test is one obvious approach. Finally, we may note one set of responses which came only from Sepang. It seems that these students were alerted to the provisional nature of their judgements and advised to use the word "possibly". A number of them did. However, rather than writing something like "the relationship is possibly a moderate, negative, linear relationship", in several cases the students transferred "possibly" to another position which led to statements like "the relationship is a moderate, possibly negative, linear relationship", even though the inverse nature of the relationship is in no doubt. Character based languages like Chinese do not provide as many clues within their construction about the relationship between the ideas embedded in each character. It is not possible to tell whether the students concerned had a problem with English or a genuine misunderstanding. However, it is possible to note this issue, and to pay special attention to it in future with any students for whom English is a second language.

Question 4 (a) - Interpretation of r^2

Since the *interpretation* of r^2 is not affected by the initial value of r all consequentially correct answers were accepted here. Using this criterion, most students did the calculation correctly (88%, 99%). Of this large difference in what is essentially a button pushing exercise, 4% is explained by Australian students who did not obtain any value for r in the first part of the question and a further 4% who obtained a value for r but did not attempt to calculate r^2.

Inadequate attention was paid to some aspects of this question when collating the Adelaide data so a direct comparison is not possible here. Only 58% of the correct respondents stated that r^2 showed that 32% of the variation of Y was explained by either X, the variation of X, or the regression (all three forms were accepted as correct for the purposes of the examination). The Sepang responses were much more uniform. Of the correct respondents 76% believed that r^2 showed that 32% of the variation of Y was explained by one of these three features—X (15%), variation in X (59%), the regression (2%). The Sepang responses suggested significant rote learning.

The Australians provided a wide variety of other responses: 4% believed that r^2 measured the percentage of explained *change,* 23% that it measured the proportion of the *time* that concurrent change occurred, 6% that it measured the probability that the variation could be explained. In contrast, there were very few other responses from Sepang—the most common was concurrent change (3%). However, 10% gave no response at all. The concept of the coefficient of determination is not easy, especially when, as in this course, it is not presented as the ratio of the explained sum of squares to the total sum of squares. It is not clear why there were quite different profiles of "other" responses. But even though the Sepang responses suggests a stronger element of rote learning than the Adelaide ones, the significant number who gave no response suggests just as strongly as the very varied Australian responses that this concept is generally not well understood.

Questions 4 (c) and (e) - Interpretation (and Reservations) of Predictions of Intercept and Slope

While part of this question is concerned merely with calculations, it also provides some indication of the extent to which the students perceive regression to be a deterministic, rather than a stochastic, concept. Research with undergraduate New Zealand students has found them to have "a tendency towards overly deterministic thinking", which "reflects their lack of awareness or understanding of variation" (Pfannkuch & Brown, 1996) and some Spanish students believe that association is a deterministic concept (Batanero et al., 1996, p. 166).

The standard form of the regression equation used in the course is $\hat{Y}_i = b_1 + b_2X_i$ where $\hat{Y}_i$ indicates the estimated value of Y_i. For this question the correct results were $\hat{Y}_i = 2096 - 129\ X_i$ where the units are in dollars. The calculations were usually done correctly (85%, 84%). There were several places where students could indicate their understanding of the stochastic nature of the regression, and these will now be discussed.

One measure of whether students appreciated that this equation is a stochastic equation can be gained by seeing whether they indicated, by using either the "cap" or words, that this equation provides an estimate of Y_i. There were weaknesses with this approach because some of those who did not include the estimate idea in their original equation did use it at other stages in their interpretation, and this was not recorded during the analysis. However, the percentages which did include the estimate idea at this stage (60%, 70%) suggest, at the least, that the significance of the symbols used in the equation has been overlooked by a significant minority of both classes.

In interpreting the coefficients of the regression equation some students did use "estimate language", and not all of this was recorded. However, one form was recorded which is particularly interesting. Some students used the term "on average", as in "the selling price of a motor bike will decrease by $129 per annum on average". This is of course reasonable. But of the 21 Australians who used the term for one parameter only, 16 applied it to the slope, 5 to the y-intercept. Of the 11 Malaysian students who used the term for one parameter only, 5 applied it to the slope, 6 to the y-intercept. These results are strikingly different. One might hypothesis that the slope, which extended over a long time period was more likely to be seen as an average than the y-intercept, so the Australian figures are not surprising, but it is not clear why the Sepang results are so close. The strongest differences between the groups arose in discussing their reservations about the regression. Of those who correctly calculated the regression, most (63%, 85%) were well aware of the dangers of extrapolation, and indicated in some way that the estimate for six months (question 4 (d) (ii)) was outside the range of the data. (It is true that an expression of reservation about the y-intercept was not asked for, but no-one who was concerned about extrapolation to 6 months expressed any concern about extrapolation to 0 months.) The difference in percentage is quite strong. The Sepang data suggests that this important point was strongly emphasised and heard by the students. But it was strongly emphasised in the Australian course as well. It is not clear why so many Australian students did not hear it.

But the Australian students were far more willing to express other reservations. Relatively many more expressed concern about the low coefficient of determination (45%, 28%) or the small value of the sample (17%, 7%). The examination paper stated that there were three marks for this part of the question, which strongly suggested that there were at least three reasonable grounds for having reservations. So all of these results are disappointing, especially the Malaysian ones. It is possible that the traditional way of dividing questions into clear parts, which makes marking very much easier, discourages students from seeing the context as a whole. It would be interesting to see what differences arose if different members of the same class were presented with two different versions of the same question—one balkanised, the other holistic.

But on another matter the results were even more disappointing still. Very few students (9%, 5%) questioned the assumption of linearity. It is difficult to believe that so many young adults are unaware of the distinctly non-linear nature of prices for used motor vehicles. Nor did students comment on the unrealistically low value of the motor bikes. Freudenthal (1991) and Schoenfeld (1988) have argued that students merely play the teacher's game, conforming as closely as they can to what they believe the rules to be, regardless of "common sense". They hold to this position even when they are strongly encouraged to question the reality of the problems they had been set.

It is possible that gender is relevant to this issue. Of the 11 students who questioned the reasonableness of this question, 10 were male, and the one female was a mature age student with substantial self-confidence, and a particularly determined approach to her studies. Ability does not seem to be relevant: the final percentages of those who had reservations were fairly evenly spread between 29 and 94.

For the Malaysians there were cultural issues as well. What interpretation could Malaysians bring to the question when the prices were expressed in a foreign, unspecified currency? The significant minority of Malaysians who did offer other reasons suggests strongly that the other reasons and the importance of contextualisation were discussed in classes, but the students seemed unwilling to offer such answers. Unfortunately, because gender cannot be deduced from Chinese names, it is not possible to see if there was a gender bias in the responses.

Conclusion

The connoisseurship approach presented here which links exploratory data analysis, pedagogical and psychological theory, and critical analysis is able to provide a deeper understanding of the students' understanding of the concepts of association and regression. What follows is a personal appreciation of the question and answers which tries to combine systematically a value judgement with a description which might help others to view the students' understandings in a different light.

The analysis suggests that the question was straight-forward and reasonable. While some marks could be obtained by accurate calculations, and others by rote learning, others required deeper thinking, and, perhaps more importantly, an appreciation of the whole problem. It is possible that the fragmented structure of the question inhibited this overall view, and it is also possible that deeper investigation of some ideas, such as "variation" and "linear" would have exposed more misunderstandings than were apparent. Most interesting of all is the strong finding that even where there is substantial evidence of general rote learning among the class, a significant minority still do not succeed.
The analysis revealed both similarities and differences between the two groups. Some differences probably arose from linguistic difficulties experienced by the Sepang students, but others probably represent cultural differences. In general, the Sepang responses were more uniform, and less critical than the Australian ones. And it was only a small minority of Australian males who were inclined to be critical.

In terms of marks the Sepang students did better than the Adelaide ones overall. To some extent this was caused by their high computational accuracy. But this analysis shows that the situation is more complex than that indicated by marks, and identifies areas where both groups could benefit from further help. Such an approach is labour-intensive, but is efficient in isolating areas of importance. In particular, has the potential to focus teaching emphasis where it is most needed. Also, it is possible to make more deductions about process learning for the group as a whole than Lipson (1994) has suggested, though not for individuals. Given that more and more teaching and examining is likely to be done in groups for economic reasons, this is not a serious disadvantage. Finally, the responses make one question whether the traditional approach of dealing only with linear association in a first-year course is educationally sound.

Acknowledgements

I thank Ms Vivien Piovesan, University of Adelaide, for her co-operation and for providing me with access to the scripts. I am grateful to those MERGA members who contributed to the discussion when Truran (1995) was presented, especially Professor Peter Jones, and also to Mrs Anne Arnold, University of Adelaide, and Mrs Kath Truran, University of South Australia, for commenting on earlier drafts of this paper.

References

Batanero, C., Estapa, A., Godino, J. D. & Green, D. R. (1996). Intuitive strategies and preconceptions about association in contingency tables. *Journal for Research in Mathematics Education*, 27 (2), 151 - 169.

Batanero, C., Godino, J. D., Vallecillos, A., Green, D. R. & Holmes, P. (1994). Errors and difficulties in understanding elementary statistical concepts. *International Journal for Mathematics Education Science and Technology,* 25 (4), 527 - 547.

Eisner, E. W. (1985).*The Art of Educational Evaluation: A Personal View.* London: Falmer.

Franklin, L. A. (1988). Clarifying regression concepts using 3 point data sets. *Teaching Statistics,* 10 (1), 8 - 12.

Freudenthal, H. (1991). *Revisiting Mathematics Education. China Lectures.* Dordrecht, The Netherlands: Kluwer.

Goode, S.M. & Gold, E.J. (1987). Linear regression and correlation—an enlightening approach. *Teaching Statistics,* 9 (2), 60 - 62.

Hawkins, A., Jolliffe, F. & Glickman, L. (1992). *Teaching Statistical Concepts.* London: Longman.

Laviolette, M. (1994). Linear regression: the computer as a teaching tool. *Journal of Statistics Education,* 2 (2). <http:www2.ncsu.edu/ncsu/pams/stat/info/jse>.

Lipson, K. (1994). Assessing understanding in statistics. In J. Garfield (Ed.) *Research papers from the Fourth International Conference on Teaching Statistics (ICOTS 4).* Minneapolis, Minnesota: International Study Group for Research on Learning Probability and Statistics.

Pfannkuch, M. & Brown, C. (1996). Building on and challenging students' intuitions about probability: can we improve undergraduate learning? *Journal of Statistics Education,* 4 (1). <http:www2.ncsu.edu/ncsu/pams/stat/info/jse>.

Popkewitz, T. S. (1988). Institutional issues in the study of school mathematics: curriculum research. *Educational Studies in Mathematics,* 19, 221 - 249

Schoenfeld, A. (1988). Problem solving in context(s). In R. I. Charles & E. A. Silver (Eds.) *The Teaching and Assessing of Mathematical Problem Solving* (pp. 82 - 92). Reston VA: Lawrence Erlbaum, NCTM.

Shaughnessy, J. M. (1992). Research in probability and statistics: reflections and directions. In D. A. Grouws (Ed.) *Handbook of research on mathematics teaching and learning* (pp. 465 - 494). New York: Macmillan.

Skemp, R. R. (1971.) *The Psychology of Learning Mathematics.* London: Penguin.

Tamura, H. (1994). Model Comparison in Regression. *Teaching Statistics,* 16 (2), 47 - 49.

Truran J. M. & K. M. (1996). Research into the learning and teaching of stochastics. In P. Sullivan, K. Owens & B. Atweh (Eds.) *Research in mathematics education in Australasia 1992 - 1995* (pp. 345 - 364). No Place of Publication: Mathematics Education Research Group of Australasia.

Truran J. M. (1995). Some undergraduates' understanding of the meaning of a correlation coefficient. in B. Atweh, & S. Flavel (Eds.) *MERGA 18: Galtha* (pp. 524 - 529). Proceedings of the Eighteenth Annual Conference of the Mathematics Education Research Group of Australasia (MERGA) held at Northern Territory University, Darwin, Australia, July 1995.

Wilkie, D. (1980). Pictorial representation of Kendall's Rank Correlation Coefficient *Teaching Statistics,* 2 (3), 76 - 78.