EXTENDING THE USE OF STATISTICAL PACKAGES IN AN ELEMENTARY STATISTICS COURSE TO ENHANCE UNDERSTANDING

James Nicholson, Belfast Royal Academy, Northern Ireland

There is a growing literature advocating the use of in-class sampling experiments to help build statistical intuition and to make the subject matter less abstract and more directly accessible to the student. The author has developed a number of such experiments in relation to correlation and regression, different sampling methods and hypothesis testing. They are used in conjunction with class discussion, and sharing data from experiments, to explore the way randomness behaves, and to try to construct a realistic concept of variation and the difficulties it causes in the realm of statistical inference. This paper outlines in detail how one class used an investigation to explore the nature of stochastic behaviour in the context of correlation and regression, and tries to draw out some of the ways that they gained some useful insights in the process.

I have been teaching statistics as part of British year 14 university preparation in mathematics for a number of years. My background is as a mathematician with some statistics in my primary degree. Over the past few years, as the technology has improved, the curriculum has moved away from number-crunching and routine application of statistical techniques, towards interpretation of data. As I re-evaluated teaching strategies to adapt to the demands of changing curricula and assessment, I started to use more practical activities. I was aiming to build up sound conceptual frameworks by exploring how random events behave, and using class discussions to draw out different perspectives and try to reconcile them into a larger scheme.

The rationale for practical work in statistics as a general principle is well documented, for example, the Cockcroft Report (Cockroft, 1982) which notes that many concepts in statistics need time and exposure to mature. Moses (1992) advocates the use of in-class sampling experiments to help ‘build statistical intuition and to make the subject matter less abstract, less formal, and more directly apprehended by the student in dealing with concepts such as the power of a test. Garfield and DelMas (1994) identify power as a concept which has fundamental implications for statistical understanding which contrasts with its non-inclusion in many first courses in statistics - including the Associated Examining Board A-level syllabus taught in this course. Since the mathematics required for a ‘proper’ treatment is beyond the scope of such courses, power simply does not appear, understandably. However, without some appreciation of power, there is a tendency for students to pay too much attention to small data sets, and too little attention to 'less impressive' differences in large data sets. Fischbein (1990a) makes the observation that: *A main source of difficulty concerns the reconciliation of the stochastic nature of statistical phenomena with the deductive structure of the corresponding mathematical schema*, in trying to summarise difficulties encountered in probability and statistics studies. Begg (1995) provides an interesting analysis of the way that the statistics curriculum is developing in various places in the world, and argues for group activity, where interaction is encouraged, to be used as a way to help students construct their own cognitive links, and to learn to communicate appropriately. The development of my investigations, and the methods of their implementation, are designed within this framework, and aim to provide experiences of the sort that Fischbein would look to in building correct intuitions. Fischbein (1990b) argues that intuitions can be developed - that intuitions are cognitive beliefs, characterised by their apparent self-evidence. Sometimes intuitions are wrong, and such misconceptions can sometimes be overcome by confronting them with appropriate experience.

I believe that it is necessary for students to undertake some 'experiments' manually, so that they fully appreciate the process that the computer is carrying out repeatedly very quickly. However, it has become increasingly apparent that even modest statistical packages open up wonderful opportunities to experience and observe many more instances of such behaviour than one could hope to do without electronic simulations. The details of some of these can be found in Nicholson (1996a, b; 1997a, b). Rossman (1997) reports on a full course, which approaches the learning of statistical concepts through self-discovery by using technology.

In the curriculum prior to A-level, the class had had informal dealings with correlation and lines of best fit. Before starting formal correlation and regression we reviewed the informal concepts they had met previously, drawing out the ideas such as residuals, and using these to justify the least squares criterion, and then deriving the line of regression. I then used a reaction ruler with one of the pupils. This measures the time taken to respond to the ruler being dropped. There was a discussion about whether we would expect him to get the same result on subsequent attempts. This was then carried out and an informal comparison of the results against their predictions encouraged the students to speculate on what factors would be influences. Without characterising this as hypothesis formulation I encouraged them to speculate, or to articulate their preconceptions as to what a full investigation might show. The following are a few of their ideas. The class felt that there was likely to be considerable variation in individual 'times' and that averaging a number of attempts would be better than using a single time for an individual. Age was thought likely to be a factor, with performance worsening in older people. Boys were expected to do better than girls, by both boys and girls! Other factors raised included alcoholic intake, tiredness, eyesight and whether it mattered which hand was used, and a variety of possible reasons why left/right handedness might have an effect.

Discussion followed as to which factors we could investigate reasonably. It was decided not to take account of eyesight as this would require equipment/expertise we did not possess, but to measure before and after a period of exercise to see if there was a difference in performance. It was also decided to leave the alcoholic intake in the realms of speculation (teacher's veto!) and to collect 10 results for both hands for individuals, and record age, sex, and which their main hand was. The pupils would collect 15 results each, trying to have a reasonable mix of boys and girls, a variety of ages, and one or two left-handed people if possible. The individual results varied considerably. Since the students had set out to vary a number of parameters this was not surprising. For the small samples not all correlation coefficients were even significant despite the pooled results giving a correlation of 0.7, which is highly significant in a large sample. This generated discussion about how good the testing procedure was, informally looking at the concept of power. Early on, they noticed substantial variations in their regression lines, and found 'estimated' or 'predicted' values over the range 8 to 28 seconds, using the different regression lines. The underlying line of regression is evident from the line of the medians, but the increasing divergence of their predictions as the *x*-value moves away from the centre helped them to appreciate some of the sources of uncertainty in using the regression line as a predictor. Initially, most of them would have said that the line was a reasonable, but not good, predictor, since the data lay in a broad band rather than close to a straight line. However, when they considered their individual sample regression lines they began to revise their view of how uncertain the predictions would be using only one relatively small sample.

The tiredness factor provided a useful starting point for discussions which really have their origins in the importance of the design of experiments. The constraints of our timetable meant that they decided to undergo five minutes of physical exercise in the school's gymnasium, using weights or a rowing machine, and to measure the reaction times before and after exercise. The correlation coefficient for the 13 points was only 0.4 and since there was only one sample, there was a tendency to accept that this was the 'true' coefficient, despite the evidence of the wide variety of correlations in the original data collection exercise! Many of the preconceptions and misconceptions our students bring to class are deep-seated and it is important that concepts are met in a variety of contexts if we are to be sure that they have a genuine understanding. There is a subtlety here which I think is an unacknowledged source of considerable difficulty for many students, and that is the multiple levels of stochastic behaviour. The correlation coefficient is a stochastic measure, but it is also a random variable with its own sampling distribution, and the value obtained is a function of the particular set of observations, which may or may not be typical.

Debate centred initially on the flaws in recording, or actual errors. I think this was because their intuition told them tiredness must have an effect. If the data did not show this, they reasoned that there must be an error, rather than thinking in terms of how ineffective a small sample is in determining such an effect. One interesting development was when they moved from thinking of 'errors' as mistakes, and started thinking of why the results might differ from what they had predicted, and what they were still firmly convinced was true, that tiredness would increase the reaction times. Two main points arose in this: 1) the 'experiment' required the participant to be actively concentrating i.e. they knew that the ruler was going to be dropped, and the conscious concentration might overcome the tiredness. This is quite different from the sort of context they were using to construct this firm hypothesis, for example, where a car driver is drowsy after a long journey, and an unpredictable event occurs to which he or she has to react. They could see that the measured response could well be different from that which they were trying to model. And 2) there were too many confounding variables. Each pair of values was from a different person and there were definite variations in the degree of exercise undertaken. Of course, much in the real world is multivariate data, and formal introductory statistics courses can only extend to a treatment of bivariate data. However, I believe that there are dangers in restricting students' perceptions to single 'causal factors', particularly where the term 'errors' rather than 'residuals' is the common language. Hawkins *et al*. (1992) argue that some intuitive appreciation of partial correlation may be made by judicious use of graphical overlays or computer graphics to identify subgroups. In this investigation, comparative boxplots, and scattergraphs with different symbols for subgroups were used in exploring the underlying relationships.

I also learnt more about the ways that faulty interpretations and misconceptions arise from hearing them articulated in discussion than I would have from answers to any examination question, where the focus is on very specific issues. For example, one student characterised the full results group, where the correlation coefficient was 0.7, as showing that there is “no difference in reaction times of the two hands” and we were able to explore what sort of scattergraph would be expected if that were the correct conclusion.

Later in the course these students investigated a moderately large data set by using samples generated by Minitab. This was similar to the investigation reported in more detail in Nicholson (1996a) except that the underlying dataset was larger. Their interpretation of the regression line, and of the reliability of predicted values from the regression lines, was more appropriate than I have seen in previous groups. The classroom environment is crucial to the process. Over the past few years I have learnt something of how to encourage students to be open with one another. Almost everyone can contribute something worthwhile with some effort, and their contributions are useful even when they are flawed; they have a chance to address their construction of the concept and make appropriate modifications. The accessibility of computer statistical analysis means that it is increasingly important that those using it have a feeling for the concepts involved and a respect for the underlying conditions required by particular techniques. Possibly even more important is that they need to be able to communicate with the non-statistical community in a way which preserves statistical integrity, but is able to convey effectively the outcomes of the analysis.

Shaughnessy (1997) identifies variation as one of the major missed opportunities in research into the teaching and learning of data and chance. As an experienced classroom teacher, I am convinced that such activities, focused to provide a stimulating environment and properly supported by classroom discussion, can enhance students' understanding. Formal evaluation of such conceptual development is much more difficult, but I believe it presents us with a worthwhile challenge in the early part of the twenty-first century.

REFERENCES

Begg, A. (1995). *Statistics and the Mathematical Processes* Teaching. Statistics, *17(2),* 40-43.

Cockcroft, W. H. (Chairman) (1982). *Mathematics Counts*. HMSO.

Fischbein, E. (1990a). Postscript *Training Teachers to Teach Statistics Ed. Hawkins A.* ISI.

Fischbein, E. (1990b). Training Teachers for Teaching Statistics *Training Teachers to Teach Statistics Ed. Hawkins A.* ISI.

Garfield, J. and delMas, R. (1994). Students' Formal and Informal Understanding of Statistical Power in *Research Papers from ICOTS 4.*

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

Moses, L. (1992). The Reasoning of Statistical Inference. In D. C. Hoaglin and D. S. Moore (Eds.) *Perspectives on Contemporary Statistics.*. MAA Notes 21. USA. 107-122.

Nicholson, J. R. (1996a). Exploring Sampling. *Teaching Statistics, 18(2),* 49-51.

Nicholson, J. R. (1996b). Understanding Statistical Concepts, *MicroMath. 12(2),* 24-25.

Nicholson, J. R. (1997a). The Relative Effects of Sample Size and Proportion. *Teaching Statistics, 19(2),* 47-48.

Nicholson, J. R. (1997b). Developing Probabilistic and Statistical Reasoning at the Secondary Level Through the Use of Data and Technology. In J. B. Garfield and G Burrill (Eds.), *Proceedings of the 1996 Roundtable Conference, Granada* (pp 29 - 44) ISI, Voorburg, The Netherlands.

Rossman, A. (1997). Workshop Statistics: Using Technology to Promote Learning by Self-discovery. In J. B. Garfield and G Burrill (Eds.), *Proceedings of the 1996 Roundtable Conference, Granada* (pp 221 - 232). ISI, Voorburg, The Netherlands.

Shaughnessy, J. M. (1997). Missed Opportunities in Research on the Teaching and Learning of Data and Chance. In F. Biddulp and K. Carr (Eds.), *Proceedings of the Twentieth Annual Conference of the Mathematics Education Research Group of Australasia Incorporated.* K. 6 - 22.