The Pre-requisite Knowledge Required by Students Starting Data Analysis Classes at University

Susan Starkings
South Bank University, Learning & Development Centre
Borough Road
London, England
starkisa@sbu.ac.uk

Jerome Healy
London Guildhall University
London, England

Abstract

"Students, especially in the behavioural sciences, often believe that while other courses may be enjoyed, statistics courses must be endured." Erickson & Nosanchuk (1995)

Many of the students joining the Data Analysis classes at South Bank University have adopted this culture and it is often because of previous mathematical/statistical experiences. The social science students are often mathematically weak and do not respond very well to using traditional textbook techniques, however, by careful instruction these students can cope with the material being presented to them. The use of good quality software, sympathetic teaching and appropriate material can not only enhance the learning experience but also enable the students to understand the need for the statistics classes. Most tutors expect the students to have a minimum mathematical competency level. This is often not the case at South Bank University and extra sessions to bring the students up to this level of mathematical competency has been required. This paper be will threefold namely (i) to state the pre-requisite mathematical knowledge required; (ii) how this knowledge can be taught and (iii) the educational implications for other subject disciplines.

Pre-requisite Mathematical Knowledge

The tutors teaching on the Data Analysis units for the social science students at South Bank University have a great deal of problems with students who are taking these statistically based units. The tutors have identified potential reasons, from research carried out on previous students, why students have difficulties with the subject matter. One reason is that many students are uncomfortable with numerical data and lack the intuitive feel for interpreting it. This is a skill that with enough practice and exposure to various examples and data sets can be understood and put into action as and when required. A second reason given is that statistics, like all mathematical based subjects, has a jargon of its own which can confuse the non-statistician and often makes the subject look more difficult than it really is. The third and main reason identified by the tutors is the use of symbolic/mathematical notation and the student's lack of basic mathematical concepts. Basic concepts that may seem trivial to some students can inhibit the learning process of the majority of social science students. For example basic conversions from one form to another, such as 50% is 0.5 or one half is accepted by those in the know. For students who have not grasped this then the route to understanding the statistics is already at a disadvantage. It is important to realise that symbolic notations are little more than a form of short hand for concisely and accurately writing down expressions, which would otherwise be much clumsier.

During the summer of 2000 South Bank University, Learning and Development Centre gave two remedial courses in statistics to Social Science students. The students were studying for degrees in Sociology, Social Science, and Government and Social Policy. The courses consisted of three hour sessions delivered over 5 and 6 weeks respectively, to 1st and 2nd year students needing to re-sit examinations and / or re-submit coursework for the Data Analysis modules of their courses. This experience was useful in gaining insight into the key skills required by students undertaking non-specialist courses in Data Analysis and Statistics. It also stimulated reflection on how these skills might be taught to students coming from disparate educational backgrounds who were basically non-quantitatively orientated. It was discovered that the students exhibited a number of problems with mathematical attainment that increased the difficulty of the Data Analysis module for them.

- The students were extremely reliant on calculators for basic arithmetic but lacked confidence in using more advanced functions, and in particular the statistical functions, available on their calculators.

  A fellow lecturer stated the following scenario; it goes like this;

  A lecturer was explaining to a class how to use a calculator to generate (pseudo) random numbers. The lecturer says "You see, when I push the '#' button I get the random number .023". A student puts up his hand looking puzzled and says, "Excuse me, but how is it when I press the '#' button on my calculator I get a different number, .052".

  Although this similarity maybe seen as unrelated and unrealistic. However, one of the first surprises in giving the course to the 1st year students related to the use of calculators and was very similar to this scenario. Having spent some time explaining the calculations required in obtaining the variance the lecturer stated. "Once you have obtained the variance, it is a simple matter to obtain the standard deviation, since the standard deviation is just the square root of the variance". A student response was "Excuse me, but how do you find the square root".

- Many students had real difficulty with mathematical notation

  The next problem found was a difficulty with formulae, the expressions including Σ, in which statistics abounds, causing particular nervousness in many students. When a calculation was worked numerically the students could understand it. However, attempts to generalise it by using symbolic notation appeared to confuse many students. The difficulty with formulae and notation no doubt reflects inadequate prior training in algebraic manipulation, and the students showed little facility for this.

- Extended calculations appeared problematic for the students.

  A third major problem was that the students were clearly accustomed to performing only trivial arithmetic calculations and unused to performing the more extended calculations required for data analysis. The most complex calculation that 1st year students were capable of dealing with was calculation of the variance of non-frequency distributed data. The 2nd year students displayed a generally higher level of mathematical competence, nevertheless the maximally complex calculation they could manage was finding the variance of frequency distributed data.

- The use of an appropriate text book

A most important aspect of learning statistics is finding a book, which explains statistics as simply as possible. There are a large number of statistics books on the market written specifically for students who are studying the social sciences and others written specifically for students who do not have much of a background in the subject. The advice from the unit leader on this issue is to consult several texts until they find one that is pitched at their level. Students can waste a great deal of time by scanning textbooks.

The topic that caused most difficulty for both the 1st and 2nd year students was that of frequency distributions. However most of the difficulty that the students found with the material were due to an insufficient level of mathematical attainment. The problems with mathematical attainment displayed by these students mirror those described by Phoenix, D., (2000) in discussing students in the Life Sciences.

Teaching Strategies

The lectures deal with practical use of statistics. Unnecessary jargon and symbolic notation is avoided to focus more clearly on the underlying principles. Given the problems in mathematical attainment displayed by these students, as discussed above, the following strategy was adopted for the remedial courses;

- The emphasis was on inculcating an intuitive grasp of the principles rather than rigour. Due attention being paid to teaching the required calculation procedures.

- The use of symbolic notation was confined to the minimum necessary with numerically worked examples being used wherever possible.

- A step by step verbal explanation of each calculation was provided in the lecture notes, and all technical terms explained in simple language.

- All examples used related to the student's area of study at an appropriate level, to motivate the need and usefulness of statistics.

Education Implications

The experience in teaching these courses leads one to conclude that the mathematical preparation of students who intend to undertake degree courses in the Social Sciences containing non-specialist modules in Data Analysis needs to emphasise the following minimal key skill set.

- Algebraic manipulation to at least the General Certificate in Secondary Education taken by 16 year olds

- Extensive practice in the use of formulae and in substitution of numeric data into formulae and their subsequent solution for the variable of interest.

- Proficiency in extended calculations.

- Preparation of graphs and histograms from numerical data.

Students who have studied mathematics or statistics to Advanced level will generally possess these skills. However, with the expansion of higher education in the UK students now come to university by a variety of routes which may not include a high level of mathematical preparation.

The extent of the student's reliance on calculators to perform arithmetical manipulations, even when these were trivial and could be more easily performed mentally by moving decimal points, for example. At the same time, they had a weak grasp of the use of the more advanced functions offered by their calculators. This raises the question of whether use of calculators at elementary school level. This might be harmful to developing mental arithmetic skills and a feel for numbers. Indeed, some countries ban use of calculators in elementary school mathematics to help develop the basic feel for numbers.

Summary

The student's difficulties could be described succinctly as a lack of confidence in their ability to tackle the contents of the Data Analysis module attributable to a less than ideal level of mathematical competence. Gal et al (1997) stated that "Students' attitudes and beliefs can impede (or assist) learning statistics, and may affect the extent to which students will develop useful statistical thinking skills and apply what they have learned outside the classroom." The lecturer giving a remedial course such as this needs to be able to empathise with students who may have no particular interest in mathematics or statistics and may even actively dislike the subjects. These were students for whom the Data Analysis module of their course was a compulsory unit, and would not have been chosen as an option. Motivating such students presents a challenge, but it can be done given a sympathetic and imaginative approach.

REFERENCES

Erickson B H & Nosanchuk T A (1995), *Understanding Data 2nd Edition,* Open University Press, Buckingham

Gal, I, Ginsburg L & Schau C (1997), *Monitoring Attitudes and Beliefs in Statistics Education*, The Assessment Challenge in Statistics Education, Gall I & Garfiled J B (Eds), IOS Press

Phoenix, D., 2000, *'Core Zero and the Life Scientist'*, Mathematics Today, Vol. 36, No 5.

RESUME

Les difficultés que les étudiants éprouvent peuvent être décrites comment provenant d'un manque de confiance en leur abilité d'aborder le contenu du module 'Analyse des données' qui est attribuable à un faible niveau d'aptitude en mathematiques. En outre, il est essential que tout professeur donnant des cours de rattrapage soit capable de comprendre des étudiants qui n'ont aucun intérêt en mathematiques ou statistiques voir même qui éprouvent un profond dégoût pour le sujet. Ces étudiants faisaient partie du groupe pour qui le module 'Analyse des donnèes' était obligatoire et qui n'auraient pas choisi de suivre comme option. Motiver de tels étudiants représente un défi qui peut être relever en penant une approche imaginative et supportive.