Teaching Multivariate Analysis to Business Students

James Clayson - Paris, France

1. Institutional background

I would like to describe to you some of my experiences in teaching a middle level statistics course to business undergraduates at the American University of Paris (AUP). But, in order for you to see how the nature of my institution has affected my view of the role "statistics" can play in undergraduate education and how it might more effectively be taught, I must say a few words about this unique institution. After these few remarks I will locate for you both me and my statistics class within its curriculum.

The AUP is a small, four-year, American liberal arts college located in the centre of Paris. It has a full-time undergraduate population of about 1000, a full-time faculty of 35, and a part-time faculty of 70. The majority of the students are not American and are drawn from 65 different countries. English is the language of instruction, but for many students and faculty it is neither the first nor second language.

AUP offers the BA degree in art history, comparative literature, European cultural studies, international affairs, French studies, economics, and international business. A single BS degree is offered in computer science. More than 40 percent of the students are in the business programme, while computer science - the only technical degree programme in the college - has just a handful of students.

The AUP curriculum is strong in languages, humanities, and the social sciences, but weak in mathematics and science. For example, while AUP has always required that each undergraduate be proficient in French and take a wide variety of humanities and social science offerings, a laboratory science requirement was only recently installed. There is still no college-wide mathematics requirement and only the computer science, economics, and business departments require their students to take a single semester of statistics. It is not surprising then to find that most AUP students - like most American liberal arts undergraduates - have little idea of science or the scientific method, and have no interest in learning about it.

2. Role of statistics in a liberal arts programme

I arrived at AUP in 1982 to redesign the mathematics and statistics offerings. I tried to revivify the traditional "Introduction to Statistics" course by introducing a computer laboratory component, student data gathering exercises, and by using a text that was less traditional, more chatty, and more directed towards the integration of exploratory with confirmatory statistics (*Understanding Data*, Erickson and Nosanchuk). I was not satisfied with the computer software available (SPSS etc.) - I found them too powerful and too tediously difficult for my students to learn quickly, so I programmed a series of interactive implementations of basic techniques in APL. My programs were not so general but were immediately usable for the tasks I had in mind. In addition, at least for that ancient time, they produced a wide variety of simple graphics. (More recently we have begun to use the Macintosh implementation of DataDesk for the laboratory exercises.)

I entered teaching late; I was 40 when I began in earnest. Because I came to AUP from a career in operational and marketing research, I knew how statistics worked in real-world situations and I wanted to share these enthusiasms with my students. I had just attended ICOTS 1 in Sheffield and was eager to try out many of the things I had heard discussed there. But, in addition to this, I wanted to introduce the flavour and power of the scientific method to students who probably had had no taste of it elsewhere. Because of the paucity of mathematics and science offerings at AUP, each course that was offered - especially the few required courses - needed to be as effective as possible. Most likely they were the last such courses most of my students would ever take.

I had a free hand to do whatever I wanted inside the classroom but I was constrained on the material to be covered. And so, apart from exploratory data analysis, the syllabus of the course remained traditional: hypothesis testing, ANOVA, and linear regression.

My first problem - and it has remained a problem ever since - was finding the right kind of teachers for this new course. Of the six sections of this course offered per year, I was able to teach only one. For the others I had to rely upon more traditional teachers who would, when possible, sit in on my own course for a term.

3. Results

I was in a rather privileged position to view the results of this "Introduction to Statistics" course, at least for the business majors who form the majority in the class. Business students take statistics in their first or second year; in their final year they are required to take a course in operational research (OR). I teach all of the sections of the OR course so see all the business students who will be taking a degree from AUP.

I see little trace in these students of ever thinking through some of the big ideas of introductory statistics. Often they cannot remember even the simplest definitions and formulas. I like to think that my statistics students have more glimmer in their eyes than the others but this may be a question of language since I spoke continuously to my statistics classes about *big ideas* and they were used to responding to that phrase. They remembered a view of statistics as a system of thought - not merely a collection of tests. But most students could honestly remember little. They should have taken a course in tennis or golf instead; at least then their bodies would have remembered something from the time spent.

While there is much talk about researching the ways students *learn* there is less interest in the ways students *remember*. Obviously, AUP students were *learning* the material in their basic statistics class well enough to pass the required examinations and to do the laboratory work. Obviously, too, they forgot what they learned almost immediately. Why?

4. Reasons why statistics courses fail

My strong feeling was that there were at least three main reasons why basic statistics courses - like the one described here - fail to excite a greater number of students.

First, the course offered statistics as a non-related collection of techniques. Hence, at the end of the semester it was difficult for students to state - or remember - what the big ideas of statistics were. I have found no textbook that is organised around major concepts and so it is up to the instructor to impose the *big picture* on the class. Each new technique introduced must be described in relationship to all the other techniques, the notions that link them together should be described carefully in words, and students must be encouraged to formulate the concepts in their own language using a minimum of formulas. Examinations need to test ideas as well as computational abilities.

Second, the course material was not as stimulating as it might have been. When we remove much of the mathematical underpinnings of the standard introductory course, what remains is - more often than not - tedious manipulations towards no *intellectually* satisfying end. We must be careful not to remove the statistical *metaphors* and statistical *ideas* when we attempt to minimise the mathematical content.

Third - and this may sound a bit Californian - much of the course content is not personalised enough for the students; they cannot become emotionally involved with the material. Data sets need to be drawn from situations that are important and relevant to students.

With these thoughts in mind, I decided to try a different approach to teaching an introduction to statistics course. I would not attempt to replace the traditional offering, but would offer an elective course in which I could experiment with a different syllabus. I could then compare these students with others to see whether the altered approach had a noticeable impact.

5. Behaviour research - a different approach

When I was studying statistics at university I found that many of the best ideas were saved until graduate school or at least until a fair amount of advanced mathematics was ploughed through. This was fine for someone with the dedication, but what about the student who either could not get through the mathematics or who saw no reason for doing so? Why not, I thought, reverse the order of topics and introduce some of the exciting notions of multivariate analysis immediately. Certainly the vision of cutting down dimensionality through principal component analysis, for example, might be a more exciting starting point to students than the lengthy approach to hypothesis testing that introduces traditional statistics courses.

I decided to group a number of multivariate techniques that especially interested me and were, I believed, all related in ways that would make for useful class discussion. In addition, since this was to be an elective in the business programme, I wanted to discuss techniques that were appropriate to current practices in market and economic research organisations, and that were implemented effectively in software packages that my students could use on the Macintoshes available at AUP's computer lab.

Finally, I named the new course *Behaviour Research* to emphasise my intention of gathering and analysing data that were pertinent to the lives and behaviour of the students attending AUP. I selected the following techniques: multiple regression, principal component analysis, cluster analysis including minimal spanning trees, multi-dimensional scaling, questionnaire design and contingency tables. For the first several classes I used a number of books from the Sage series: *Quantitative Applications in the Social Sciences*. More recently I rely on my own notes for the class.

I would like to illustrate my approach in this class by talking a bit about one topic that incorporates most of my themes.

The American psychologist George Kelly argued that each of us tries to make sense of our world with techniques similar to those used by scientists. Kelly gives the name *constructs* to the patterns that people are continually testing against reality. Constructs "are ways of construing the world. They are what enables man to chart a course of behaviour, explicitly formulated or implicitly acted out, verbally expressed or utterly inarticulate, consistent with other courses of behaviour or inconsistent with them, intellectually reasoned or vegetatively sensed" (Kelly, 1963).

Each person's scientist aspect encourages him to improve his constructs. For Kelly, *behaviour* is *experiment*. Rather than merely responding to surroundings, people use an experimental approach to test and extend their system of personal constructs to better explain and predict personal environments. Kelly believed that one's construct system could be improved by making its elements verbally explicit and his *repertory grid technique* aids individuals towards this goal.

A few terms: Kelly's constructs are bipolar scales that measure similarities and differences. Each end of a scale is considered to be the opposite of the other and in-between measurements will be made in reference to both ends. Obviously, different number scales can be investigated, but I have found a 1-5 scale to be useful.

The things being construed are called elements. Elements can be anything from, say, a list of automobiles to members of one's family. Typical people-construing constructs, for example, might be: tall/short, fat/thin, sensitive/non-sensitive. Usually the elements are first defined by the person playing the Kelly game and then some method is applied to elicit constructs from the player. Each construct is used to evaluate or construe each of the elements. Eventually, a *repertory grid* will be formed with the columns representing the elements, and the rows the constructs. The numbers within the grid will be the evaluations of the elements across the constructs.

Kelly's method for eliciting constructs selects three elements at random and asks the player: "What construct would make two of these elements alike and different from the third?" The idea is to encourage the player to suggest constructs quickly without too much thinking. After all, the method was devised to make visible constructs that the player had never consciously viewed before. The goal is to *trick out* an individual's constructs as fast and as easily as possible and I programmed a software package in APL to elicit constructs in this way.

I have found that *thinking about one's family* is a good first grid exercise. Few people have trouble visualising family members in their mind's eye and, although they may never have construed them in the way I suggest, they are eager to give it a try. Here is an opportunity, too, to show students how statistical approaches can encourage them to think more deeply and effectively about their own personal life.

Kelly stressed the importance of including a wide variety of characters in such an exercise. Obviously, if we are to elicit constructs that differentiate between elements, we must offer a wide variety of elements. Here is the list of my suggestions that I give my students for their family collection.

1.	Mother	Biological, step, or adopted by you
2.	Father	Biological, step, or adopted by you
3.	Siblings	Favourite or least favourite
4.	Grandparents	Biological, step, or adopted by you
5.	Uncles/aunts/cousins	Biological, step, or adopted by you
6.	Priest/rabbi/minister	Real or ideal
7.	Teacher	Best, worst, ideal, rejected or rejector
8.	Supernatural figure	Real or imagined
9.	Historical figure	Famous or infamous
10.	Real me	As you know yourself to be
11.	Other's me	As others see you
12.	Wished me	As you wish to be
13.	Friend	Current, ideal, old, rejected or rejector
14.	Flame	Current, ideal, old, rejected or rejector
15.	Enemy	Current or former
16.	Hero	Real, historical or ideal

Once a grid is obtained, the students naturally want to find out what constructs are the most alike, where the redundancies are in one's construct system, are there surprising similarities between constructs that might link, say, physical characteristics with moral or emotional judgements, are some constructs totally independent of most others, ...? And what about the similarities and dissimilarities of individual family members; who, for example, is most like *me*?

At this stage several statistical idea-methods for comparing matrix rows and columns are suggested by the students and my APL package implements some of the comparison notions of distance and variance.

Various distance metrics can be selected although the city block metric is the default. This leads into a single linkage cluster module that generates easily-read graphics. The graphical displays show the original grid data, its reordering by rows and columns in order to place the most similar rows next to each other and the most similar columns next to each other. This moving around of rows and columns gives a visual guide of the effect of the distance metric/cluster method being used. The Mac then draws the trees. In addition, the program structure is fairly interactive so if a student wishes to see the effect of a change in any of the original data, he makes the changes and immediately sees the results in terms of changes to the tree shapes. (For a description of cluster methods applied to repertory grids, see Shaw (1980).)

Next, the grid matrix is transposed so that the constructs become the columns and the elements become the rows. This array is then decomposed into its principal components. If there is enough redundancy in the set of constructs - that is, high correlations - to make this exercise useful, the elements are plotted in principal component space; and the student can select whatever dimensions he wants. Because my program does not yet allow for dynamic rotation, I introduce perspective by changing the size of points in a two-dimensional plot according to the third dimension.

To indicate clusters I calculate a minimal spanning tree in whatever principal component space is selected. The program then lets the user step through the tree, starting with the shortest links. The linked item names are listed on the screen as the link is drawn on the diagram.

Finally, the correlations between the principal components and the original columns - the constructs - can be plotted. This chart shows graphically the blend each principal component is of the original constructs and this encourages the student to make sense of the meanings of the principal components.

This principal component aspect of my repertory grid program is also fairly interactive. That is, the student can watch the effect any change in the original data will have on the patterns found in the graphical PC output.

6. Results

I have found this single exercise to be very effective in reinforcing some of the big ideas of not just multivariate analysis, but statistical analysis itself. It encourages the student to explore a personally interesting problem using a collection of complementary techniques. I have only anecdotal evidence that statistical ideas verbalised in this fashion are more likely to stay in a student's head than if introduced more traditionally. But my skimpy evidence is all positive. In addition, I have solid documentation that students who take my Behaviour Research course are able to get good jobs using these techniques. Why? They were able to describe multivariate techniques in *words* to the interviewers they had to face.

In summary, I have found that when statistics is introduced and explored as a system of interlocking ideas and methods, and using data that is relevant to student's lives, the likelihood that these students will *learn* and *remember* individual techniques is enormously increased.

References

Kelly, George (1963) *A Theory of Personality : The Psychology of Personal Constructs.* W W Norton, New York.

Shaw, Mildred L G (1980) *On Becoming a Personal Scientist : Interactive Computer Elicitation of Personal Models of the World.* Academic Press, London.