

SUCCESSFUL STRATEGIES FOR TEACHING MULTIVARIATE STATISTICS

Joseph F. Hair, Jr.
Kennesaw State University, United States
jhair3@comcast.net

Perhaps among the most challenging course I have taught is multivariate data analysis. The topic not only is difficult, but most students are ill prepared to take the course. In such situations, one must first overcome the initial fear factor of students and then get them engaged enough to keep them motivated. To accomplish this, I utilize the following teaching strategies: (1) represent yourself as a real person to your students; (2) use lectures sparingly; (3) encourage active learning through discussion and hands-on exercises; (4) include lots of examples; and (5) constantly reinforcement previous concepts. These represent the strategies I have found to be the most successful.

INTRODUCTION

In my years as a professor, I have taught many different courses. Perhaps among the most challenging courses I have taught is multivariate data analysis. Indeed, the course represents a challenge for both the professor and the students. This paper discusses how a teacher can not only teach the research skills needed by students, but also introduce them to a broad range of complicated, yet very useful multivariate statistical techniques. The course begins with an overview of research design, sampling, measurement, questionnaire design, and data collection approaches. It then provides an overview of the primary multivariate statistical techniques, discusses examining your data to ensure its accuracy and completeness, and then focuses on in depth, hands on coverage of the major techniques.

BACKGROUND

My interest in multivariate data analysis began some 35 years ago with a quite shocking mandate from three new, young professors who had just been hired in the marketing department of the business college at the University of Florida where I was ABD (all but dissertation) in my doctoral program. Social science research and particularly marketing were changing rapidly. Quantitative approaches were emerging along side traditional qualitative methods such as case studies. For my cohort of six doctoral candidates, the shock was the requirement that our doctoral research must use multivariate data analysis. The problem was that none of us had taken a course in multivariate statistics. Indeed, there were no courses on the topic in the college of business and only limited offerings in the math/statistics department, with the emphasis on solving equations, not application of the techniques. Thus began a journey that led me to write the first edition of my text *Multivariate Data Analysis: With Readings* (Hair et al., 1979).

All of the books on multivariate statistics at that time included many formulas and/or much matrix algebra. For example, the text *Multivariate Data Analysis* by Cooley and Lohnes (1971) noted the authors were not statisticians, but rather data analysts. It also suggested that multivariate statistics held the keys that would unlock the secrets of human social behavior. But for students with limited math or statistics backgrounds, this and other texts were still a huge challenge to read and comprehend. The applied statistical concepts were simply overwhelmed by the formulas and math! The other popular text of that time, *Multivariate Statistical Methods* (Morrison, 1967), claimed to be written as “an elementary source for multivariate techniques” but was far from elementary for virtually all business college doctoral students.

The initial challenge for my cohort of doctoral friends was to learn enough multivariate statistics to complete the doctoral program. This task opened my eyes to the potential of multivariate techniques, but clearly identified the need for a more user-friendly approach to teaching them. After much frustration in seeking a relatively readable source to learn multivariate, I concluded there was none. My goal thus became to write a text that had no formulas and emphasized what the techniques could do, how to select the correct method to achieve a particular research objective, and how to properly apply and interpret each technique. We almost achieved that in the first edition – there was only one formula, for the concept of a variate – the linear

combination of multiple variables and a basic building block of all multivariate statistical techniques. We also described the major multivariate statistical techniques and their application in plain English and provided an example of how to interpret the results. Perhaps one of the most innovative pedagogical techniques of the text was the use of HATCO (Hair-Anderson-Tatham Company) as a single database throughout the book to demonstrate the application and interpretation of each technique.

The real challenge of that first edition (1979) actually was not writing the book, but getting it published. At that time my coauthors and I were unknown scholars from less prestigious universities located in the southern part of the USA. Moreover, the field was viewed as a niche area with little potential. It took six years and a lot of persuading to find someone to publish the book, and even then it was a small publisher new to the area and willing to take a risk. The first edition sold about 3,000 copies, enough to make it successful for the publisher. With the expansion of interest in the field, the text was later acquired by a major publisher MacMillan and eventually by Prentice-Hall. The applied approach of the book along with improved, more user-friendly software and PC technology have all combined to make the book the leading text in the field today by a wide margin.

Writing the book was just a first step, however, in learning the best approach to teaching the topic. Knowledge of a field is only about one-half of the basis for success in teaching. The other half is the ability to communicate that knowledge. In the following paragraphs I share with you what I have learned about effectively communicating the concepts of multivariate data analysis to students.

REPRESENT YOURSELF AS A REAL PERSON TO STUDENTS

Very early in the course you must establish yourself as a real person – someone your students can relate to and feel comfortable with. This is particularly true if you happen to be an author of the text used in the course. How is this done? Sometimes it can be done with humor. For example, begin the first class with a question like “What kind of person would choose to spend their time writing a text on a topic like multivariate data analysis? They must be pretty weird!” I then continue the discussion by telling the story of how and why I wrote the book to demonstrate the rationale, and to show them how an academic’s career and personal life can reinforce each other.

Another way to establish your self as a real person is to be approachable. Since with the multivariate course students often are older and have business experience, it is not necessary to insist they refer to you as “Dr.” in order to achieve respect. They will respect you for how you teach and help them to learn, not by any titles you may have earned. You can also be perceived as approachable by spending time with students outside of formal classes. This may mean having a drink with them or sharing personal experiences where you have common interests. But it also provides an opportunity to informally share your knowledge of the field. In short, make sure you show empathy, understanding and a willingness to go the extra mile to see that they learn the subject.

USE LECTURES SPARINGLY

My lectures are based on my text *Multivariate Data Analysis* (2006). I ask students to read the appropriate chapters before class. Indeed, I caution them that just reading the chapters once is not sufficient. It usually takes two to three readings of the chapters before they begin to grasp the material. I also urge them to study the definitions of key terms at the beginning of the chapters. Multivariate is in some ways like a foreign language. After all, “Do any of you consider words like homoscedasticity or orthogonal to be part of your regular, everyday vocabulary?”

I refer to my presentations as lectures, but in a typical session my comments represent only about one-third of the time. My objective is to provide an overview of the major issues and then move on to a more interactive approach. The remaining two-thirds of the presentations involve answering questions, group work and active learning. *PowerPoint* slides are used to support my lectures and be a resource for follow-up applications of the techniques. My goal is to distill the chapter material down to the essence. The details can be picked up with the fourth and fifth reading of the chapters, or through in-class discussion and questioning.

ENCOURAGE ACTIVE LEARNING

Learning is not a spectator sport. Students do not learn by just sitting in class listening to teachers, memorizing prepackaged assignments, and spitting out answers. They must talk about what they are learning, relate it to past experiences, and in the case of multivariate data analysis, actually apply the techniques. One strategy I use to facilitate active learning I refer to as a “Learning Checkpoint.” This involves showing a slide with a series of questions covering the major concepts I expect the students to have mastered at that point. I ask students to mentally rehearse their answers to the questions, and then they form groups and discuss and agree upon the answers. Depending upon the concept, I sometimes just ask students if they are confident they know the correct answer, while in other instances I actually have them give brief verbal answers to the entire class.

Another strategy involves having students actually apply the techniques. This requires teaching either in a computer lab setting, or having students bring laptops. It also means the appropriate software must be available – in most instances, I use *SPSS* but *SAS* is also a possibility. For the most part, your selection of software depends upon what is available and you are familiar with as there are several excellent alternatives. As a further method of facilitating application of the techniques, I have prepared a database that gives logical, believable results. For example, in a recent book I published, *Essentials of Business Research Methods* (2003), I created the Samouel’s and Gino’s case study. Samouel’s is a Greek restaurant that competes with Gino’s Italian Ristorante. All students have eaten out at restaurants and can relate to the situation. There are two databases – one is a survey of customers’ perceptions of the two restaurants and the other is a survey of employees of Samouel’s Greek Cuisine. I typically show students how to apply the techniques with one of the databases and then have them replicate my examples on their own computers. Then, I make similar assignments for the other database and ask them to apply the appropriate techniques to solve the problem. Rather than telling them which variables to use, the learning starts with asking them to review the database, identify the appropriate variables and statistical techniques, and then use *SPSS* to run the solution. The results in are summarized and presented to the class, often using a group format. While the students are completing this assignment, I circulate throughout the class to clarify questions and ensure students are moving in the right direction.

Since many of the students do not have extensive experience with *SPSS*, I prepare a handout with the appropriate click-thru sequence for each of the techniques. This handout provides simple instructions on how to use *SPSS* to execute statistical analysis and leaves time for more important topics like how to interpret the results.

The value of a single database to teach all the techniques is something I learned teaching with the HATCO case study through the first five editions of my multivariate book. But after more than 20 years the field has witnessed many advances, not only in the user-friendly software and inexpensive hardware, but also in the field of research methodology. One major advance has been the field of measurement. This develop has been supported by the more widespread familiarity with multivariate statistics which facilitates examination of multiple variables simultaneously. As a result, we recognized a need to update the HATCO database and the 6th edition of the multivariate book (Hair, 2006) features a new and expanded database – HBAT. The HBAT database is still a business-to-business case study, but now includes more relevant issues and most constructs are multi-item measures. Moreover, in the chapters on structural equations modeling (SEM has three chapters for more thorough coverage) the example application can be executed from start to finish with the raw data, and includes coverage of multiple group analysis, item parceling, longitudinal data, and interpretational confounding, to mention a few.

EXAMPLES, EXAMPLES, EXAMPLES

Teachers almost never give enough examples. This is particularly true in multivariate data analysis. But providing examples is just the beginning. The examples must be relevant and supportive of the concepts. For example, how do I communicate the concept of an outlier? I obtain real data on Bill Gates as an example of an outlier (see Table 1). I use Bill Gates because he is well known, his information makes the point well, and accurate information is actually available.

Table 1: Impact of Outliers

<u>Individual</u>	<u>Net Worth</u>	<u>Source</u>
Bill Gates	\$46.0 billion	Microsoft
Jeff Bezos	\$5.1 billion	Amazon.com
Craig McCaw	\$2.0 billion	Telecommunications
<hr/>		
Number of households, Medina (Seattle), Washington	=	1206
Average Net Worth (1206 households)	=	\$ 44,253,482
Average Net Worth (remove Bill Gates)	=	\$ 6,115,934
Average Net Worth (remove top three)	=	\$ 224,189

Source: Calculated from *Forbes*, 2003 and U.S. Census data

Another concept important in teaching multivariate is the necessity of assessing error. After defining error and listing the sources, I give concrete examples. For example, to demonstrate respondent error I talk about how I have heard my wife complete telephone surveys on several occasions. When she comes to the question on annual household income she always responds but she really has little idea what our annual household income actually is. This not only provides a good example of respondent error, it also supports the earlier strategy of portraying yourself and your family as real people. Another interactive approach to teaching respondent error is to ask students how many credit cards they carry in their purse or wallet? When asked the question, very few students check their wallets; they simply reply. I have used this exercise for years and the responses given by students are consistently wrong about one-third of the time. A third way to provide a good example of how errors are introduced into research studies is to refer to studies in the literature that have documented error. One of my favorite is the classic study on alcohol consumption in which respondents reported one level of consumption (the socially acceptable response) to interviewers, but when their actual consumption was verified through inspection of their garbage it demonstrated a significantly higher consumption level.

When teaching the individual statistical techniques, I also include situations in which I have actually applied the techniques. Sometimes my examples are scholarly and based on published articles in the literature. Other times they are from previous consulting experiences. These applications range from the use of multidimensional scaling to develop perceptual maps for a food service company, to discriminant analysis to facilitate the location strategy of a financial services firm, and conjoint analysis to identify the preferred product for a software company. Such examples bring to life the techniques because they clearly demonstrate their application in a context that is meaningful to students.

REINFORCEMENT

Knowledge is retained through reinforcement of the concepts. Because of the level of difficulty of the topic of multivariate data analysis, I encourage students to attend my classes more than one time. Some of the classes I teach on the topic are one-week workshops. I typically teach these mini-workshops several times a year, most often at universities in Europe, but also in the Pacific Rim. The best approach is for students to attend first as an initial exposure to the possibilities when they are identifying a topic for their doctoral research. Then they attend again when they are planning their data collection to reinforce the earlier sessions and to refresh their knowledge of research design issues such as scaling, construct development and questionnaire design. Finally, they attend after their data is collected because then they can apply the techniques to their own data, which is truly an opportunity for active learning. For semester-long courses like those typical in the US, I still encourage students to attend more than once and they often do. As the old saying goes “practice makes perfect” and this is definitely true with multivariate statistics.

CONCLUSION

Knowledge of multivariate data analysis is clearly important for graduate students and scholars today. But it is increasingly important for the average individual. Data mining using multivariate statistical techniques impacts almost everyone today, from the tax authorities using discriminant analysis to identify individuals to audit, to both brick-and-mortar and online companies using it to evaluate new product ideas and develop more effective marketing strategies.

In addition to the strategies outlined above, I have found the following observations to be particularly applicable not only when teaching multivariate data analysis, but with other topics as well:

- Courses need to have practical relevance and provide students an opportunity to understand how the concepts they are learning can be applied in the real world.
- Courses need to add value to students' careers by contributing to their knowledge, skills and values.
- Students will be successful if you communicate to them what needs to be done to succeed, and how to do it.
- Your presentation approach is just as important as the content being taught. Presentations should include pauses for emphasis, animation, enthusiasm, visuals, story-telling, and changes in your tone of voice.
- Finally, learning often is hard work, but it can be enjoyable, positive and even fun.

REFERENCES

- Cooley, W. and P. Lohnes (1971). *Multivariate Data Analysis*. New York: Wiley.
- Hair, J., Anderson, R., and Tatham, R. (1979). *Multivariate Data Analysis: With Readings*. Tulsa, Oklahoma: PPC Books.
- Hair, J., Babin, B., Money, A., and Samouel, P. (2003). *Essentials of Business Research Methods*. New York: Wiley.
- Hair, J., Black, B. Babin, B., Anderson, R. and Tatham, R. (2006). *Multivariate Data Analysis* (6th edition). Upper Saddle River, NJ: Prentice-Hall.
- Morrison, D. (1967). *Multivariate Statistical Methods*. New York: McGraw-Hill.