COMBINING STATISTICS AND INFORMATION SYSTEMS: THE LA SALLE EXPERIENCE

by Joseph R. Troxell
Department of Management
La Salle University
Philadelphia, PA 19141

ABSTRACT: In attempting to meet accreditation standards, La Salle University combines a second semester statistics course with an introductory course in information systems, as part of a two course sequence in statistical information systems. Of importance is finding ways to integrate the material from these separate disciplines. Instructors begin to do this integration in the first semester of this two-course sequence by emphasizing two points. First, the statistical data found in the text are presented as part of a data base, not just as columns of numbers. Second, a statistics computer package used in the course is used to demonstrate features of a management information system (MIS), and is not just a tool for data analysis. This article discusses these two points from a first semester perspective as an integrative way of combining material from two separate courses. In the second semester, data bases and management information systems are fully discussed as topics in their own right. Another method used in the second semester to integrate the two courses is the use of class projects, with details of one project discussed.

Keywords: Statistics, Management Information Systems (MIS), American Assembly of Collegiate Schools of Business (AACSB)

INTRODUCTION

In 1987, the School of Business at La Salle University modified its curriculum in response to two pressures: to increase the quantitative background of students in upper division courses and to bring our curriculum in line with the American Assembly of Collegiate Schools of Business (AACSB) accreditation standards mandating coverage of information systems. Unfortunately it was only possible to add one course to the required curriculum; thus was born "statistical information systems", QNT215. All Sophomores in the School of Business are required to take the resulting two semester sequence course in statistics and information systems; the first semester covers topics in statistics, while the second semester combines material from statistics with material from an introductory course in information systems. Effective integration of this material, which is not obviously related, is a challenging task. Unless the material is effectively combined, the students may find the course as much of a joke as the following course listing:

> *"SPRING BULLETIN: **Yeats and Hygiene, A comparative Study:** The poetry of William Butler Yeats is analyzed against a background of proper dental care. (Course open to a limited number of students.)" (1)*

As additional colleges and universities apply for AACSB accreditation, they too might find a course of this sort necessary. This article discusses the approach we arrived at for teaching QNT215 and identifies two topics for integrating these two disciplines, databases and management information systems.

DIFFICULTIES IN DESIGN OF THE COURSE

When considering either a statistics course or an information systems course, too narrow a view overlooks ways to combine these two courses. In statistics, data are a collection of numbers arising from some planned experiment performed under some set of conditions. It is not common to think of statistical data as part of a database, but rather as an isolated set of figures. First, the many fields typical in a database do not exist with statistical data. A typical text example or homework problem consists of an explanatory sentence and the raw data. Next, statistical packages are not presented to the students

as database oriented, although they are. The user does not expect to perform many of the database commands, such as find, sort and prepare reports; instead the package is used to merely grind out the summary statistics.

On the other hand, in an MIS, statistical methods are just one of many kinds of applications that can make up an application base. Other application bases might contain methods suitable for analyzing problems in marketing, production, accounting, and so on. In a transaction processing system, little use of statistical applications is made; in a decision support system, statistical methods are useful for some unstructured decision making tasks, but not all. In an information systems course, spending any length of time on a statistical method, for example, regression, does not make sense. Little of the flavor of information systems carries through.

To achieve the necessary integration, stress must be placed on thinking of a set of statistical data as a database. Once the concepts of databases and database management are discussed, the analogy with statistical data is straightforward. Then, a statistics package can be presented as satisfying the definition of a management information system. After discussing management information systems, their relation to a statistics package is also straightforward.

Of importance is recognizing that integration begins in the first semester of the two course sequence. The first semester of this course covers the statistical topics of descriptive statistics, probability, probability distributions, inference, confidence intervals, hypothesis testing, and simple linear regression. In any given semester, four or more faculty teach sections of this course, each having different backgrounds in information systems ranging from little to extensive. For that reason, the first semester course content is primarily statistics. However, instructors need to make two departures from the usual statistics course content. Doing this in the first semester makes the second semester material more integrated. In addition, it emphasizes several things about statistics and data that need to be covered, and that most textbooks omit.

STATISTICAL DATA AS A DATABASE

First, the instructor emphasizes that data come from experimental units, each of which is distinct and has distinguishing characteristics. Modern data plotting techniques emphasize the use of color, symbol coding and the like to reveal differences in data due to differences in the experimental units. The use of statistical data to help managers solve problems and make better decisions is framed around the well-known I-P-O model of inputs, a process, and outputs. Here the instructor first defines the model, then emphasizes how using statistical methods transforms meaningless data into meaningful information.

> ***Our course seems even more threatening (than David Kroenke's "Widow-maker" Introductory Management Information Systems course); perhaps we should call it the road to early retirement.***

Generally, statistical texts treat data as isolated sets of numbers, with little background information. This includes 'famous' sets of data that have been analyzed in many ways, often to present new tools of analysis. Frequently the readers of these datasets have little idea of what the data represent and how they were collected. The instructor also emphasizes the question of why the data are collected. While many texts offer a wide variety of interesting sets of data, the larger question of why the student or researcher wants to examine the figures is ignored. On the output side, he/she emphasizes the use of a report to present the analyzed data in a coherent way. Many texts simply ask students to calculate various statistical measures, but seldom ask them to write even several sentences to explain what is learned by examining the figures.

In information systems, a database is the last piece of a hierarchy that progresses bit, byte, field, record, file and database. Database Management Systems focus on record processing which is also called file processing. Thinking of a record as a group of related fields, one wants to include all relevant fields for a complete record. Database Management Systems give users a variety of field descriptors, so that users can designate a field as alphanumeric, numeric, a date, monetary, and so on. Then too, indexing features allow for key field designation so that physically distinct files can be linked. A record can be one field to any number of fields in length.

When described as part of a database, statistical data are now viewed in new ways: as raw data to be included in a report for management decision making; as one variable in a multivariable database, including descriptors of the experimental units that were sampled; and as numbers which can be edited by sorting, finding specific subsets, and further processing. By beginning a discussion of databases in the first semester, students feel comfortable with them when they use database management products in the second semester.

STATISTICS PACKAGE AS A MANAGEMENT INFORMATION SYSTEM

At this point, an MIS is defined as consisting of an application base, a model base, a database management system (DBMS), and a database existing on storage medium (2). A statistics package can also be described as a management information system. The description need only be brief, with detailed discussion of management information systems as part of a hierarchy of systems left for the second semester. A brief description begins by explaining that the user of an

MIS wants a collection of applications to be available, with the ability to design others on an ad-hoc basis. Applications typically involve the generation of reports, including word processing and graphics. The user also wants a collection of models for analyzing the data; in statistics the models are the statistical methods such as regression, the analysis of variance, and Chi-Square tests.

The instructor emphasizes that the user of a DBMS must learn three languages found in a DBMS; a data definition language, a data manipulation language, and a high-level query language. He/she also emphasizes that companies seldom want a single database for all their corporate data. With new relational databases available, integration becomes easier, but firms still use databases for separate functional areas of marketing, production, accounting, and so on.

At La Salle, we use Exploring Statistics with the IBM PC (3) as our undergraduate statistics package; it demonstrates the application base, model base and database management features described in the brief outline of an MIS above. A main menu offers three submenus consisting of a set of 24 programs, a file manager, and a help program describing all of its menu choices. The application and model base consist of fifteen statistical methods, such as simple data description and multiple regression. (The package also contains five programs particularly relevant to the teaching of Statistics, such as a program to simulate the Central Limit Theorem.) There is no report writer or word processor included with the package, however, it does contain a program to generate simple plots of data for two variables only.

The DBMS data definition language is covered by use of the File submenu that allows the student user to create, edit, save, retrieve, or delete files. Each file contains the data on one variable only; the appropriate variables for an application are collected at the time the application is run. Other statistical packages, notably Statgraphics, are variable-oriented like this. The data manipulation language can be invoked in the program and allows for data transformations such as logarithms, and the creation of new variables through multiplication or division of current variables. No separate query language is found in the package.

The package comes with a data disk containing data from 1980 on the 50 United States; 44 variables are included. The instructor explains that the states data is a socio-economic database similar to those that social scientists use, although in practice the real database is longitudinal.

SECOND SEMESTER

When the integrative work discussed above is done in the first semester, the material in QNT215 is more coherent. Nevertheless, some awkward transitions occur and some difficult choices need to be made concerning topics and the time spent on them. No one formula is 'right', and individual instructors vary course content to satisfy their own preferences. Our collective choice is to spend six weeks on the statistics material, with the remaining eight weeks in a 14 week Semester on the information systems material. No single text book exists that combines material for a two semester course of this sort, so students buy the statistics text the first semester and the information systems text the second; in QNT215 both books are used. The second semester examinations in this course are fairly sophisticated and use problem solving, multiple choice questions, and examining computer printouts.

STUDENT PROJECTS

During the first semester the students are exposed to using a personal computer with the statistics package to perform simulations and calculations, and to generate output in the form of tables and graphs. In QNT215, the students perform three projects as part of their grade. The first project, in multiple regression, allows for the students to go through the stages of model building using a small number of variables. The emphasis is on using statistical measures as guides to building a correct model.

The second project uses dBase III to expose students to considerations in designing and manipulating databases. The student does not use the data definition language, but instead uses a database already created by the instructor. Emphasis is on using the database to answer managerial questions.

Finally, the third project uses the statistics package again and is concerned with regression model building and variable transformation. During student evaluation, equal weight is given to the student arriving at a good model and in explaining how the model building is performed using the package as an information system. Using the database that comes with the statistics package, the students are able to recreate a study done in the mid 1970's on factors affecting traffic deaths in the United States (4). The result of this study is that other factors can be used to explain the variation in traffic deaths from state to state, notably the location of the state in the country. This project is written as a scenario involving a United States Congress person who wants to make a position statement on whether to change the current speed limit law from 55 miles per hour. It is explained that an office staff person would perform the actual analysis using an even more extensive database than supplied by the package.

CONCLUSION

David Kroenke (5) calls the introductory Management Information Systems course the Widow-maker, meaning that it is a tough course to teach. Our course seems even more threatening; perhaps we should call it the road to early retirement. Nevertheless, our two semester course meets the challenge of integrating material from statistics and information systems, and makes new insights into the link between these disciplines.

ACKNOWLEDGEMENTS

I wish to thank Joseph Seltzer and the three referees for helpful suggestions in revising the paper.

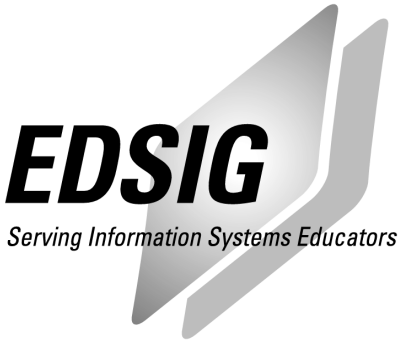
REFERENCES

1. Allen,Woody. Getting Even. Warner Books, Inc. New York, 1972.
2 Kroeber, Donald W. and Hugh J. Watson. Computer-Based Information Systems, 2nd ed. Macmillan Publishing Co., New York, 1987.
3. Doane, David P. Exploring Statistics with the IBM PC, 2nd ed. Addison-Wesley, Reading, MA, 1988.
4. Tufte, Edward R. Data Analysis for Politics and Policy. Prentice-Hall, Englewood Cliffs, NJ, 1974.
5. Kroenke, David M. "The MIS Course: Remaking the Widow Maker." Interface 10, no. 3 (1988): 2-12.

AUTHOR'S BIOGRAPHY

Joe Troxell is an Associate Professor in the Management Department at La Salle University, where he teaches statistics, quality control, and information systems. He consults to industry and pursues research in all three of these areas. While he earned his doctorate in Statistics at Rutgers University, in 1984 he attended the AACSB Institute in Minnesota to prepare for teaching courses in Information Systems.

Information Systems & Computing
Academic Professionals



STATEMENT OF PEER REVIEW INTEGRITY

All papers published in the Journal of Information Systems Education have undergone rigorous peer review. This includes an initial editor screening and double-blind refereeing by three or more expert referees.

ISSN 1055-3096