DATA ANALYSIS FOR MANAGERS

Harry V. Roberts
Graduate School of Business
University of Chicago

1. Introduction

Since about 1970 I have substantially changed my approach to teaching of basic business statistics courses:

1. Although I try to convey important theoretical ideas, data analysis comes first and takes higher priority.

2. In selecting tools for data analysis for inclusion, I weigh heavily the needs of managers.

3. The potential value of statistics is brought out in the context of applications.

4. Individual student projects displace assigned problems and examinations.

5. There is intensive reliance on hands-on, interactive statistical computing.

I believe that the result has been substantial improvement and that the trend is still positive. Students do better and more satisfying work in class, their course evaluations are more favorable (comparing well with ratings in other, traditionally more popular, courses), and feedback from former students suggests that their exposure to statistics has had more lasting effects.

2. Problems with Traditional Teaching

Basic statistics courses – in business schools and elsewhere – are seldom popular, and the reasons lie deeper than teaching methods.

1. Students sense (correctly, in my judgment) that many of the topics covered in these courses are not very helpful to managers. For example, there is usually heavy emphasis on testing of sharp null hypotheses, yet – except for their role in diagnostic checking of statistical models – these tests are rarely applicable to business decisions. (The selection of the best of two or more processes does not entail a test of a sharp null hypothesis.)

2. Few of the statistical ideas and tools most needed in business practice are covered In these courses. For example, there is little or no treatment of statistical process control, time-series analysis, experimental

design, or probability sampling. There is little emphasis on the communications skills needed to convey statistical findings to top managers.

3. As W. Edwards Deming and others have stressed, statistics – when coupled with good human relations and organizational practice – is an important key to good management. But to most students, statistics seems dull and mathematically forbidding. To overcome these perceptions, I believe that teachers must demonstrate how statistics can be used to solve important and interesting problems.

4. Statistics is rarely used widely or well by managers; there are few testimonials from businessmen about the value of statistics.

3. Important Statistical Developments

Several major developments in statistics since the Second World War have influenced my change of approach to teaching:

1. The development of probability sampling.

2. The emergence of ideas of Bayesian decision theory that help to formalize thinking about business decisions.

3. Developments in time-series analysis that have made it possible easily to extract the information needed by management and to avoid widespread popular misinterpretations of time-series data.

4. The increasing focus of experimental design and analysis on problems of business decision making.

5. Development of tools for coming closer to sound causal inferences from non-experimental data.

6. Interactive statistical computing, with emphasis on analytical graphics, beginning with early minicomputers and now increasingly flourishing in the world of the IBM PC and compatibles.

7. The modern thrust in quality and productivity, in which statistics is seen as a major contributor to improved management practice.

4. Individual Student Projects

To take advantage of these developments, one must recognize that, while most statistics professors like statistics for its own sake most students become interested in statistics mainly if the subject promises to do useful things for them. I believe that even the seemingly limited goal of developing "intelligent consumers of statistics" is best attained if students try to produce statistics on a modest scale. Only then do most students seem to become sufficiently intrigued with statistics to want to learn about statistical theory.

Based on my experience, I believe that managers can, with limited consulting support from statisticians, actually use statistics effectively for a wide range of problems. The experience of Professor William Hunter in teaching experimental design by projects and the early returns from the innovative program of Brian Joiner to create "parastatisticians" in industry provide reinforcement to my belief. Further reinforcement comes from the apparent success of many Japanese companies in achieving effective use of statistical techniques at all levels down to the foreman and production worker.

With guidance from a teacher and the use of a "friendly" interactive computer package, students can do good projects based on data sets of their own choosing, often of their own collection, even in the most elementary course. My format has been to require a series of progress reports, starting early in the course, as the project is defined and developed. Each progress report receives prompt and full feedback from the instructor, including suggestions on exposition. The obvious burden on the instructor is offset, in my experience, by the welcome relief from the monotony of reading examination or problem papers; each project is different, and almost all are interesting.

To get students started, I normally ask them to find or collect a business, economic, or personal time series that would be useful to predict and understand. Careful guidance by the instructor is needed to get the student started quickly on a rewarding series. It is highly desirable not only to have access to data but also to have access to background information about the data. Current data are more interesting than historical data. Although many students cannot get this kind of access to company data, everyone has access to current macroeconomic data and everyone can collect interesting personal data, such as athletic performance, acquisition of a new skill, or physiological measurements such as weight, resting pulse rate, body temperature, or blood pressure. Anyone, too, can collect business data by direct observation, say of sales registered at the checkout counter of a supermarket or time to serve customers at a McDonald's franchise.

Since elementary descriptive tools are introduced at the outset of the course, students can begin analysis very promptly. As the course proceeds, they hand in successive extensions that make use of new concepts introduced in the course. Later they expand the scope to include a potential leading or explanatory indicator. Finally, many do a cross-sectional regression.

Variations on this approach are appropriate in different courses. In quality control, for example, it is often possible to implement simple experiments designed to improve a process. Work sampling studies can quickly indicate about how much time is spent in doing work that either does not contribute to company goals or that represents correction of problems created by poor quality.

5. Time Series as a Top Priority for Business Students

Many critical business decisions require reasoned interpretation of time-series data about the firm: sales, market share, quality of product, inventories, gross margins, profits, promptness of collection of receivables, labor turnover, safety, the price of the common stock, etc. Lurking in the background are macroeconomic time-series such as industry sales, indices of retail sales or industrial production, rates of unemployment and inflation, and the Gross National Product. Yet general business school statistics courses often pay little or no attention to the analysis and interpretation of time-series data.

To introduce time series in an elementary way, I start with the ideas of statistical process control, defined by a random process. From this start it is easy to bring in simple ideas of probability and inferences about individual observations, which are made vivid by sequence plots, scatter plots, and histograms. Process control provides also the opportunity to relate statistics to management in a context in which the connections between theory and practice are very easy to understand.

Next one can move to simple extensions of the random model, with the random walk as the first example. Differencing then comes in as a way of doing regression without explicit regression computations. Regression can be introduced explicitly in the context of autoregression, and the regression approach can be enlarged to include modeling of deterministic effects. Multivariate regression gives an introduction to ways of handling leading-indicator series that may contribute to better predictability and explanation.

The main theoretical emphasis is on the statistical model, its formulation, diagnostic checking, and application; and on the prediction of individual observations (for example, next month's sales) rather than formal inferences about parameters. Understanding the idea of modeling also provides some protection against blundering in applications, since students are always mindful of the importance of diagnostic checking of statistical models against the data being studied.

In this way, it is easy to develop the view of statistics as probabilistic description of reality, an expression that I find very suggestive of what statistics should be about. I am continually pleased to find how often simple statistical models – such as the random walk – give a reasonable approximation to reality. Even when the simple statistical models fail to describe reality adequately, the specific failings alert us to important aspects of reality that could otherwise have passed unnoticed. For example, outliers on a Shewhart chart guide the search for assignable causes of quality variation.

6. Causal Modeling in Cross-Sectional Data

Another high-priority area for business application is cross-sectional regression and tabular analysis, with emphasis on causal modeling and in-

cluding an introduction to designed experiments. I have found that cross-sectional analysis is better understood if there has first been a serious attempt to analyze and interpret time-series data. One can always check out time-series modeling by forecasting the next day, week, or month. This gives the right perspective for any statistical analysis. One is trying to describe reality that is full of uncertainty, not to play the verbal game of significance testing.

7. Data Analysis for Managers

In recent years I have moved strongly in the directions just sketched and have prepared and repeatedly revised a set of teaching materials called Data Analysis for Managers. These materials are problem oriented, with many illustrations keyed to output from statistical packages. Informal theory is always in the background.

Students learn the statistical packages almost without effort by following the scripts used in analyses of the numerous data sets that bring out the main statistical ideas and tools.

The most serious problem with this approach has been the linkage of teaching materials to a single statistical package which, in basic courses, must be easy to learn. If that package is not widely available, students may not be able to follow up after graduation, and any one set of teaching materials may be usable only at a limited number of schools. The current or prospective availability of most packages on the IBM PC and compatibles promises to overcome this problem.

8. Recommendation

I do not urge other teachers to follow my particular strategy, although I would of course be happy if some choose to do so. I do most strongly urge reexamination of coverage of basic teaching with the aim of giving more prominent emphasis to tools useful in actual business practice and to instruction in the proper use of these tools. In this way, business students can glimpse immediately what statistics has to offer, learn more from the course, and further their subsequent careers in business.