Teaching Biostatistics to Postgraduate Students in Public Health

Peter A Lachenbruch - Los Angeles, California, USA

1. Introduction

This paper describes how biostatistics is taught in US Schools of Public Health. The audiences of the courses are students who are *not* biostatistics majors. They want to learn biostatistics because they perceive (I hope, correctly) that the methods can help them in their graduate studies and careers. Some of these students may have a minor subject in biostatistics as part of their doctoral programme. These courses usually do not require calculus as a prerequisite. I wrote to the Departments of Biostatistics at US Schools of Public Health to get information on their programmes. I did not use a questionnaire, but in my letter I noted that I was interested in the content of their courses and how they used computers in their courses.

2. The first course in biostatistics

The audience for the first courses is first year Public Health students. Schools of Public Health must require a biostatistics course for their MPH degree to be accredited by the Commission on Education in Public Health (CEPH). Most of these students have not had any mathematics beyond high-school algebra, and that was often 5 or 10 years ago. Their mathematical skills are frequently rusty. "Math phobia" is prevalent in these courses. Faculty members in all departments often work on arithmetic skills in discussion sections and office hours. A related problem for schools that use computers in these courses is that this is often the first time the students have made a computer do something useful. Some schools use the biostatistics sequence to introduce the students to computing.

In many schools the percentage of students whose native language is not English ranges from 10% to 50%. We often assign these students into a course that strains their mathematics ability when they are also struggling with oral English. They survive, much to their credit, but the amount some comprehend is a problem.

In the United States some universities operate on a quarter system (three quarters per year), but most use the semester system (two semesters per year). The CEPH biostatistics requirement is "one course". In schools that operate on the quarter system, there are 40 contact hours in such a course. In schools that operate on the semester system, there are 45 contact hours. The five hour difference in contact time is not serious. All schools offer full year sequences. The coverage of this course is the full year sequence (three quarters, or two semesters). These courses cover the contents of most elementary biostatistics texts. Differences among the courses are greatest in the first semester or quarter because the different schools break at different points. After a year, they are all close to the same coverage.

In comparing the responses, some course descriptions were much more detailed than others. Different schools may cover the same topics, but not include the topic in the course description. Nevertheless, it seems that all programmes contain material on:

Descriptive Statistics - central tendency, spread, histograms, scatter plots.
Probability and Distributions - independence, mutually exclusive events, Bayes' theorem, binomial distribution, Poisson distribution, normal distribution.
Estimation - point estimates, confidence interval estimates.
Hypothesis testing - one and two sample tests for normal observations with variance known and unknown, paired observations, tests for equality of proportions.
Contingency tables - chi-square distribution, test of independence, homogeneity, Fisher's exact test.
Regression - simple linear and multiple regression, selection of variables.
Analysis of variance - one factor designs. Some consider multiway designs.

The schools do not agree on including correlation, regression diagnostics, multiple comparisons, analysis of covariance and non-parametric methods, although these are extremely useful to most statistical practice. Only one school mentioned sample size computation in its writeup or syllabus. Similarly, only one school mentioned Graphical Methods and Exploratory Data Analysis, although several schools teach these topics. Topics that have little coverage include Statistical Epidemiology and Vital Statistics, Bioassay, Generalised Linear Models, Bioassay and Experimental Design. Other courses taken in later years cover these subjects. No elementary courses covered Bayesian Inference.

In my first year classes, I include the topics listed above, regression diagnostics (residuals, leverage), multiple comparisons (Bonferroni, Scheffe', LSD), graphics (scatter plots, stem-and-leaf, boxplots, normal plots), analysis of covariance (noting the need to check for parallelism), and non-parametrics (rank-sum, signed rank, sign tests). I omit tests for equality of variance, and confidence intervals for the variance. Fisher's exact test gets very little time as do variable selection methods. The most important message to give the students is that it is important to think about their data and decide what to do.

When the school divides the courses over several quarters or semesters, the first semester or quarter ends at or near the unit on contingency tables. Some schools include regression in the first (and sometimes terminal) course; others place it in the second course. The first course often has a large enrolment (150 to 200 students). Opinions vary on whether we should teach these courses as one large lecture group with many discussion and lab sections, or teach them as five to eight lecture/lab groups.

Eight of the nine schools include computing as part of their introductory courses. Some schools use the campus mainframe to run MINITAB, SAS, SPSS, or BMDP. Others use microcomputers to run the PC version of these programs or SYSTAT, or another microcomputer-based package. One school responded "This year we have experimented with using a software package called STATA in the elementary course. It has turned out to be a huge success. We are going to adopt it for all our elementary courses beginning next year."

At least one school expects their students to learn to use computers. The Department of Biostatistics gives that introduction in the elementary course. Several years ago, this was an introduction to the mainframe using SAS or BMDP, but now it's done on microcomputers.

3. Other courses

Schools of Public Health offer many biostatistics courses for non-majors. For space reasons I shall only give a brief summary of the main categories, based on the responses to my enquiries. I shall be glad to provide details of the course titles and the texts used to anyone who writes in for them.

The audience for these courses consists of: (a) those taking a biostatistics minor; (b) those satisfying a departmental requirement; (c) those who want to learn about a particular method. The students minoring in biostatistics usually have a strong background in mathematics. The courses available to the non-majors fall in several groups:

(i) Public Health Statistics: Vital Statistics, Demography, Sampling Methods.
(ii) Medical Statistics: Statistical Epidemiology, Rates and Ratios, Clinical Trials, Case-Control Studies, Longitudinal Studies.
(iii) Computing: Statistical Computing, Health Data Management, Statistical Packages, Microcomputers, Simulation.
(iv) Multivariate Analysis: Logistic Regression, GLIM, Log-linear Models.

Minor field concentrations require that the student complete additional graduate level courses in biostatistics. The requirements vary, but usually give the equivalent statistical training of an MPH student.

While the courses are open to any students, some courses have prerequisites that make it difficult for non-majors to enrol.

4. Teaching methods

The traditional teaching format includes lectures sometimes accompanied by laboratories and discussion sections. The laboratories teach the students to compute the statistical procedures. Computing tools are: (a) pocket calculators (supplied by the students); (b) centralised computing facilities; and (c) microcomputers. At some schools, students can use a microcomputer laboratory for their homework. The level of integration of these microcomputer laboratories into the first year courses isn't clear; no schools mentioned them. Are these labs simply there for the students who know how to use them or is microcomputer instruction part of the curriculum? In a telephone survey of some programmes, about half the programmes have a computer lab as part of the course, and half do not. Experience indicates that integrating a computer lab into the curriculum enhances the first course. For most of the past two decades, schools used mainframe computers for statistical computation using BMDP, SAS, or SPSS. One problem was that the student either had to learn a Job Control Language (JCL), or use control statements by rote. In many courses, the instructor would pass out the necessary JCL and the control statements for the package. Very little learning of the package occurred. As students progressed, they gradually learned more of the package's control language. Statisticians still should train students in how to use mainframes, for they may analyse large problems. But larger problems can be handled on micros more easily than in the past. The vast majority of problems can be analysed on a microcomputer. Microcomputer facilities are being used for elementary and advanced courses with considerable success. A typical lab has several networked computers. This limits the size of the discussion/lab sections. The benefits of computing laboratories include developing skills that help the students in other courses; ridding them of "computer phobia"; allowing them to concentrate on the scientific problem instead of the arithmetic problem.

Other teaching methods include videotapes, case studies, audiovisual presentations, programmed learning modules, term projects and using the computer in the classroom. In the early 1960s, J S Hunter "starred" in a series of lectures shown on public television in the United States. These never reached a wide audience, and were never released to statistics departments. Video recording technology did not exist then (at least for non-professionals). Recently, the Annenberg CPB project has released a series of videotapes, *Against All Odds*, coordinated with Moore and McCabe's new book. We plan to try this in discussion sections in our introductory course in 1990. Jim Boen (personal communication) has used video recording to help teach consulting skills.

Slide-tape presentations have been available for several years. However, the examples may not be from the application area of interest; they may be dated; they constrain the instructor. About 20 years ago, programmed instruction was popular. I tried it with several topics in elementary courses, but they never were as successful (for me) as an office hour with the students.

The case study method has students critique articles in their field. Some courses have used this, especially in medical schools. Related to this is the term project. Here, the student obtains (or is given) and analyses a data set from their subject field. This generates much enthusiasm in the students, and is an effective teaching tool. But the instructor does have the problem of evaluating many different projects. Colleagues I have spoken to have liked the results.

Microcomputers can be connected to a projector that displays the results of an analysis. This is especially useful for plotting data. You can make stem-and-leaf plots, boxplots and normal plots quickly, and you can show the students the effects of different distributions. One instructor generates observations from a standard normal, a log-normal, and several mixtures of normals with equal means and different variances, different means and equal variances, and displays the relevant plots to the class. Regression diagnostic plots show how they can be used in model criticism more effectively than reading a list of the leverages or Cook distances.

5. Discussion

There is considerable similarity in the course content of the introductory courses offered to non-majors in Schools of Public Health. This is hardly surprising since there is consensus about the elements of biostatistics, and the introductory texts cover much the same material. The schools differ in the amount of biostatistics they require. The minimum is one quarter, but other schools require two quarters. My bias is for coverage of topics including contingency tables, regression, and one-way analysis of variance. A one quarter course cannot cover this much.

Schools offer many courses beyond the first. These include courses in computing, vital statistics and demography, sampling, epidemiologic topics, multivariate analysis topics, logistic regression, and log-linear models. The content and availability of these courses vary depending on the school and faculty interests.

The methods used for teaching always include lectures and laboratories/ discussions. It was surprising that computing was often not introduced until well into the course (sometimes not until the second quarter course). This is unfortunate since students can compute many elementary statistics and graphics with an increase in understanding of their meaning. Few instructors seem to use the computer in the classroom itself. The availability of the projection devices has been a limiting factor. Until recently, a projector was expensive and bulky. Now, the cost is about one-third of its previous level, and the bulk is that of an overhead projector. It is important to rehearse the presentation, and to have an alternative plan available if the computer or projector malfunctions. Videotapes are not widely used. Until recently, no tapes were available. Another limiting factor is the equipment. Large screen video projectors or sets are expensive. The projectors require precise alignment. Unless a school dedicates one classroom for video projection, the device must be a large screen set with a VCR, mounted on a cart.

In the future, we may anticipate greater use of computers in elementary courses, both as calculating tools and in the classroom as a teaching tool. Excellent statistical packages, tailored for interactive computing, are available, and will be improved. The special topic courses use computers as needed. The jury is out regarding videotapes. They require additional effort and enforce a structure on the course that is not the instructor's. The schools agree on the content of the introductory course and it will be unlikely to change much. Modifications of emphasis (e.g. graphical methods, model checking) will occur. The texts will need to change to accommodate this.