COMBINING THEORY AND PRACTICE IN A FIRST COURSE ON SURVEY METHODS

William D. Kalsbeek, University of North Carolina

1. Introduction

Since 1980 two committees of the ASA Section on Statistical Education have studied the problem of training statisticians for industry (ASA Committee on Training of Statisticians for Industry, 1980) and government (ASA Committee on Training of Statisticians for Government, 1982). A common recommendation of committees like these is that a suitable balance between theory and practice must be found for introductory courses in Statistics. Understanding the practical side of statistics requires exposure to real-life problems, ideally as hands-on experience or otherwise by some sort of large-scale application (Mosteller, 1980). Thus, learning statistics becomes a matter of both understanding and applying techniques.

The focus of this paper is this important marriage between theory and practical application as it relates to teaching introductory courses on survey methods. The basis for discussion will be our own experience in developing the course BIOS 164 (Introduction to Sample Survey Methods) in the Department of Biostatistics at the University of North Carolina.

2. Some Problems

Circumstances surrounding the development of BIOS 164 seem to be somewhat typical of those at other universities. Our intent was to produce a one-semester three hour service course which, for the majority of students, would be their only course in survey methods. Furthermore, those taking the course would come from both statistical and nonstatistical disciplines thus implying that learning priorities and quantitative skills would vary considerably among students. Given these circumstances, there were four things we had to consider in developing our course.

What Survey Topics to Cover

To make the course a useful learning experience for all class members we felt that it must cover both sampling topics for those more statistically inclined and nonsampling topics for those less interested in the statistical aspects of the survey method, while maintaining the interest of each group during those portions of the course when the low-interest topics are covered.

Ultimately, of course, it is beneficial for all students regardless of orientation to be exposed to both types of topics. This serves to ease their transition from school to job and the associated burden on their first employer. An investigation by an ASA Survey Methods Section Committee to Assess the Job Market for Survey Statisticians (1983) has found that this burden may be substantial. In a study of 11 organizations engaged in survey research the majority of professional staff were found to have acquired their nonstatistical survey skills (e.g., questionnaire design), on the job. Between a quarter and a third of these same staff members learned to design a sample after they began their job.

Time Constraints

With sampling theory an extraordinary amount of material must be covered in order for the student to understand how to analyze data from the complex sampling designs they are likely to encounter in practice. Most common designs cannot be understood until toward the end of the following sequence of topics: simple random sampling (SRS), stratified SRS, SRS selecton of equal-sized clusters, probability proportional to size (PPS), selection of unequal-sized clusters, and weighting of data from unequal probability sampling designs.

The fact that the best hands-on exposure in a survey methods course is a live survey makes matters even more difficult. Exposing students to bits and pieces of a realistic survey (e.g., by conducting a few interviews), is possible though not entirely satisfactory, since the student cannot fully appreciate the interrelationships between component parts of the survey. On the other hand, having the students do a more manageable though less realistic survey (e.g., to estimate the average girth of trees on campus) fails to expose them to the problems of survey measurement and organzation.

Most live surveys would exceed time limitations. Mail surveys take longer than would be available during a four month semester and personal interview surveys are often logistically difficult and costly unless done locally. That leaves the telephone survey as the best possibility.

Finding a Balance Between Theory and Application

On the one hand, we might have devoted full attention to statement and proof of the statistical properties of estimators under various sampling designs, much in the spirit of Cochran (1977). The emphasis in this approach would have been teaching students the mathematical origins of the key results.

On the other hand, we might have taught the statistical topics by stating and applying these same key results but without devoting extensive effort in proving them, more in the spirit of Sheaffer, Mendenhall and Ott (1979). Likewise, nonstatistical topics might have been covered by either discussing principles or by practical illustration alone. The respective advantages of the two approaches are obvious. Emphasis on theory tends to improve one's understanding of the origin of important statistical results, while application enhances one's understanding of how to use the results in practice.

Choosing a Textbook

In a course involving both nonsampling topics as well as the theory and application for topics related to sampling, a single course textbook must have several qualities. It must devote significant attention to all aspects of the survey method. Sections on nonsampling topics must examine the topics as well as consider some of the operational problems. In the sections on sampling a derivation of the key results must be given, the implications of the results discussed, and the operational aspects of their application to real-world situations illustrated.

So far as we can tell, no existing survey methods textbook meets all requirements although Moser and Kalton (1977) comes the closest. Several make especially strong contributions to certain requirements (e.g., Kish, 1965; Mendenhall, Sheaffer and Ott, 1979; Cochran, 1977; Hansen, Hurwitz, and Madow, 1953, Vols. I and II; Cassel, Sarndal and Wretman, 1977; Warwick and Lininger, 1975; and Zarkovich, 1966.

3. One Approach

Having mentioned some of the problems encountered in designing our introductory survey methods course, we now present a description of BIOS 164 as we have offered it once a year during the spring semester (January through April) since 1981. The 30-35 students that the course attracts per semester have been distributed approximately as follows between 1981 and 1985:

Description	Percent
(1) Graduate Student in Biostatistics or Statistics	55
(2) Undergraduate Student in Biostatistics or Statistics	25
(3) Graduate Student in Public Health but not Biostatistics	5
(4) Other Graduate Students	15
	100

Most graduate students are first year students in a masters degree program, and most undergraduates are seniors.

The course has three class meetings per week: a one-hour lecture on Monday and Friday and a two-hour laboratory session on Wednesday. Each class session addresses one or more of the course's three major components:

(1) Sampling theory;
(2) Principles of nonsampling survey tasks including questionnaire design, data collection, data processing, and report preparation, and
(3) A survey project done to provide hands-on exposure to sampling and nonsampling tasks in survey research.

Components (1) and (2) are considered mainly during the Monday and Friday sessions, while component (3) is the subject of most Wednesday sessions. The BIOS 164 format is illustrated in Table 1 using the 1985 course schedule.

Rather than relying on a single textbook, a typed set of notes is used as the basis for presentation during lecture sessions. The notes on sampling theory draw heavily from Cochran (1977), on sampling application from Kish (1965), and on nonsampling methods from Moser and Kalton (1971) and Warwick and Lininger (1975). Ten homework problem sets, consisting altogether of 25-30 exercises, are assigned periodically during the course. Exercises are either of three types: proofs, computional application, or empirical studies. The latter type of exercises is done in two laboratory sessions.

A live "telephone survey project," however, is a novel feature and perhaps the most demanding aspect of the course. It is done entirely by students for the purpose of meeting real objectives imposed by a real sponsor with real interest in the survey's findings. Sample sizes have varied between 280 and 320.

The key to adding a survey project to BIOS 164 was local sponsorship. The last four surveys have been sponsored by North Carolina Memorial Hospital, a large university teaching hospital located within walking distance of the classroom. The first survey was sponsored by the local county health department, located about 10 miles from the campus.

Without a sponsor to provide many of the needed physical and fiscal resources it is hard to imagine that doing each survey would have been possible. The sponsor in each survey agrees to arrange and pay for the following:

(1) Work space for interviews;
(2) Telephone installation and toll charges;
(3) Copying expenses; and
(4) The salary of a graduate student hired to coordinate training and data collection.

On the other hand, the following five activities are the responsibility of the class:

(1) Questionnaire development;
(2) Sampling;
(3) Training and data collection;
(4) Data processing; and
(5) Analysis.

These tasks become the basis for dividing up the students into what we call task groups. To reflect differences in workload among groups, the number of assigned students has, respectively, been 4, 4, 13, 4, and 5 in a 30-student class. Except for the data collection group, one student volunteer is chosen to act as coordinator for each group. The paid graduate student, one who has previously had the course, serves as the coordinator for training and data collection.

The work done by each task group is identified in advance of the course by means of detailed list of survey tasks prepared by the instructor. The use of activity lists is similar conceptually to the "task-analyses" approach to instruction suggested by Watts (1981). The list of activities establishes the scope of responsibility for the group and helps to differentiate its work from other groups.

Since students could not participate directly in all task groups, some effort has been made to expose everyone in class to work done in groups other than his/her own. First, each group conducts a Wednesday lab session in which the group's assigned activities and its progress to date are reviewed and the problems encountered are discussed. The second opportunity for indirect exposure to other groups' activities is through written reports prepared by each group at the end of the course. The reports are made required reading and become a basis for some of the questions in the final examination.

Integrating a survey project in amongst the usual lecture/lab format of an introductory course like BIOS 164 seems to require three things. First, work on the project must begin as soon as possible. We see from Table 1 that in 1985 the session to organize the project happened one week after the first class session. A second important step is to avoid getting behind schedule by setting and meeting a series of deadline dates. The set of deadlines established for the 1985 BIOS 164 survey (see Table 2) made the questionnaire development group the busiest at the beginning of the project and the analysis group most active at the end. The sampling, training and data collection, and data processing groups were busiest during the middle portion of the project. A third key to integrating the project project into the course is to logically schedule the remainder of the course around the survey so that relevant lecture topics are covered before the intensive periods of work for the groups begin.

Data from each BIOS 164 survey have been analyzed following procedures that are typically applied in practice. Raw sample weights are adjusted for nonresponse using a weighting class adjustment procedure, and then analyzed using a program called SESUDAAN (Shah, 1981) which computes appropriate estimates and variances.

In addition to meeting the sponsor's objectives, students have been encouraged to piggyback methodological studies onto the class surveys. For example, in the 1983 survey two alterative methods for choosing one adult at random in sample household were tested and compared. Although hypothesized differences in response rates under the two methods were not found, evidence seemed to indicate that the use of these randomization techniques might have been "fudged" by the interviewers. These studies have not been novel contributions to survey research but they have enabled the students to resolve for themselves some of the issues raised in lectures.

4. Evaluation

We conclude by attempting a measure of self-evaluation based on our experience with BIOS 164.

What do the students say?

In general Figure 1 indicates that students have given the course favorable ratings except for a rather disturbing downward trend in the past two years. This trend may point to a need for greater variety in the course. In addition to adding new material to the lectures this change might best be made by doing different kinds of surveys each year, although doing so would require finding a new sponsor or convincing an existing sponsor to try something different.

Are two courses needed?

Suspicions that perhaps the existing course tries to do too much leads us to wonder whether a preferred alternative to the present single course would be a two-course sequence where a first course on the theory of survey methods is followed by a second course in which the survey is done. Having a separate course to do the survey could leave more time during analysis to better understand the evaluation/interpretation step in the three step process of statistical reasoning (Chervany et al., 1977). The first two steps, comprehension and planning/ execution, seem to be adequately covered under the single course option. One problem with a two-course sequence is that the courses might best be taken right after each other, which in scheduling courses for already-crowded degree degree programs would be difficult.

How could the computer aid in teaching sampling theory?

In the present course we devote parts of two lectures and two laboratory sessions to illustrate the following important concepts: random variables, sampling distribution, expected value, variance of an estimator, and the effect of sample size on the variance of estimated means. Beyond the formal definition given in a lecture we have found that empirical demonstrations of these things are quite effective. Perhaps the simple but somewhat naive hand-calculated illustration we currently use to teach these definitions and concepts might be replaced by a demonstration using available software in which hundreds of realistic samples from specified populations can be easily selected and results presented.

To summarize, this paper has discussed the difficulties of and one approach to teaching a one-semester introductory course on sample survey methods. The course we have taught for the past five years encompasses both Cochran-level sampling theory and the principles of nonsampling survey methods, as well as a live survey project to provide students with a realistic glimpse of survey practice. Our major underlying concern with this approach is the scope of the course which currently demands a heavy time commitment from students. If as Cervantes has put it, "experience is the universal mother of science," then we must strive to find ways to more efficiently teach survey methods courses which combine theory with practice.

REFERENCES

1. American Statistical Association (1980 "Preparing Statisticians for Careers in Industry: Report of the ASA Section on Statistical Education Committee on Training Statisticians for Industry," The American Statistician, Vol. 34, No. 2, pp. 65-75.

2. American Statistical Association (1982) "Preparing Statisticians for Careers in the Federal Government: Report of the ASA Section on Statistical Education Committee on Training of Statisticians for Government," The American Statistician, Vol. 36, No. 2, pp. 69-81.

3. American Statistical Association (1983) "1982 Case-Study of Job Skills Required by Survey Researchers: Report of Findings," ASA Section on Survey Research Methods Committee to Assess the Job Market for Survey Statisticians, July 7, 1983.

4. Cassel, C.M., Sarndal, C.E., and Wretman, J.H. (1977) Foundations of Inference in Survey Sampling, John Wiley and Sons, Inc., New York.

5. Chervany, N.L., Collier, R.O., Fienberg, S.E., Johnson, P.E., and Neter, J. (1977) "A Framework for the Development of Measurement Instruments for Evaluating the Introductory Statistics Course," The American Statistician, Vol. 31, No. 1, pp. 17-23.

6. Cochran, W.G. (1977) Sampling Techniques, Third Edition, John Wiley and Sons, Inc., New York.

7. Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953) Sample Survey Methods and Theory, Vols. I and II, John Wiley and Sons, Inc., New York.

8. Kish, L. (1965) Survey Sampling, John Wiley and Sons, Inc., New York.

9. Moser, C.A. and Kalton, G. (1971) Survey Methods in Social Investigation, Second Edition, Basic Books, Inc., New York.

10. Mosteller, F. (1980) "Classroom and Platform Performance," The American Statistician Vol. 34, No. 1, pp. 11-16.

11. Scheaffer, R.L., Mendenhall, W., and Ott, L. (1979) Elementary Survey Sampling, Duxbury Press, North Scituate, Massachusetts.

12. Shah, B.V. (1981) "SESUDAAN: Standard Errors Program for Computing of Standardized Rates from Sample Survey Data, RTI/5250/00-01S, Research Triangle Institute, Research Triangle Park, North Carolina.

13. Warwick, D.P. and Lininger, C.A. (1975) The Sample Survey: Theory and Practice McGraw-Hill, Inc., New York.

14. Watts, D.G. (1981) "A Task-Analysis Approach to Designing a Regression Analysis Course," The American Statistician, Vol. 35, No. 2, pp. 77-84

15. Zarkovich, S.S. (1966) Quality of Statistical Data, Food and Agricultural Organization, United Nations, New York.

Table 1. 1985 Schedule of Class Sessions for BIOS 164

Session	Day*	Date	Topic	Component Covered		
				Sampling Theory	Nonsampling Principles	Survey Project
1	W	1/09	Course Organization;Survey Components			
2	F	1/11	Questionnaire Design 1		X	
3	M	1/14	QD2		X	
4	W	1/16	Survey Project:Objectives & Organization			X
5	F	1/18	Definitions and Basic Results 1	X		
6	M	1/21	DBR2	X		
7	W	1/23	Survey Project:Questionnaire Design			X
8	F	1/25	Simple Random Sampling 1	X		
9	M	1/28	SRS2	X		
10	W	1/30	Lab 1:Simple Random Sampling	X		
11	F	2/01	SRS 3	X		
12	M	2/04	Data Collection		X	
13	W	2/06	Survey Project:Training/Data Collection			
14	F	2/08	Equal-Sized Cluster Sampling 1	X		
15	M	2/11	ESCS2	X		
16	W	2/13	Survey Project:Data Processing			X
17	F	2/15	Unequal-Sized Cluster Sampling 1	X		
18	M	2/18	USCS 2	X		
19	W	2/20	Survey Project:Sampling			X
20	F	2/22	Sample Size Determination 1	X		
21	M	2/25	SSD2	X		
22	W	2/27	Mid-term Examination			
23	F	3/01	Data Processing		X	
24	M	3/04	Spring Break:No class			
25	W	3/06	Spring Break:No class			
26	F	3/08	Spring Break:No class			
27	M	3/11	Stratified Sampling 1			
28	W	3/13	Survey Project:Interviewer Training			
29	F	3/15	SS2	X		
30	M	3/18	SS3	X		
31	W	3/20	Lab 2:Sratified Sampling	X		
32	F	3/22	SS4	X		
33	M	3/25	Systematic Sampling	X		
34	W	3/27	Analysis and Report Preparation	X	X	
35	F	3/29	Miscellaneous Sampling Topics 1	X		
36	M	4/01	MST 2	X		
37	W	4/03	Case-study:Survey Design	X	X	
38	F	4/05	MST 3	X		
39	M	4/08	Easter Holiday:No Class			X
40	W	4/10	Survey Project:Final Edit Check			
41	F	4/12	Area Sampling	X		
42	M	4/15	Sources of Survey Error 1		X	
43	W	4/17	Survey Project:Analysis Theory			X
44	F	4/19	Survey Project:Analysis Findings			X
45	M	4/22	SSE2		X	
46	W	4/24	Course Wrap-up			

*M=Monday } Lecture for 1 hour
F=Friday }
W=Wednesday Laboratory for 2 hours

Table 2. Major Project Deadlines for the 1985 BIOS 164 Survey

Date	Activity (Group Affected)
1/21	Questionnaire first draft completed (QD)
1/22	Pretest of questionnaire first draft completed (QD)
1/23	Questionnaire design work session (QD)
1/30	Questionnaire final draft completed (QD)
2/06	Data collection work session (TDC)
2/11	Sample size and selection method finalized (S)
2/13	Data processing work session (DP)
2/20	Layout for analysis data file produced (DP)
2/20	Sampling design work session (S)
2/22	Sampling frame edited and ready for sample selection (S)
3/01	Sample selected (S)
3/01	Interviewer training manual completed (TDC)
3/12	Schedule of interviewing assignments completed (TDC)
3/13	Interviewer training session (TDC)
3/15	First draft of analysis protocol prepared (A)
3/29	Final set of analysis tables identified (A)
4/05	All interviewing completed (TDC)
4/10	Final edit work session (DP)
4/12	Program set-up specifications for SESUDAAN completed (A)
4/12	Analysis workfiles prepared (DP)
4/17	Analysis work session: theory (A)
4/19	Analysis work session: findings (A)

Key: QD = Questionnaire Development
S = Sampling
TDC = Training and Data Collection
DP = Data Processing
A = Analysis

Figure 1. Course Evaluations by BIOS 164 Students: 1981-1985