Journal of Official Statistics, Vol. 26, No. 3, 2010, pp. 395–409

# The 2009 Morris Hansen Lecture: The Care, Feeding, and Training of Survey Statisticians

*Sharon L. Lohr*[1]

The two volumes of *Sample Survey Methods and Theory* by Hansen, Hurwitz, and Madow (1953) have had great influence on the training and practice of survey statisticians. We examine current themes in survey sampling research and relate them to topics taught in classes on survey sampling. We discuss other aspects of university training and background that may help the survey statistician thrive in and adapt to a variety of environments.

*Key words:* Communication; history of sampling; probability sampling; statistical education; statistical thinking; survey sampling.

## 1. Introduction

It is a great honor to be asked to give the Morris Hansen lecture on the topic of training survey statisticians. Morris Hansen, through his books, papers, and work at the U.S. Census Bureau and Westat, exerted tremendous influence on the training of survey statisticians as well as on the practice of survey statistics. Throughout his career, he emphasized the importance of involving statisticians in solving problems and designing systems that would improve the quality of estimates. His own work was remarkable in that respect; he emphasized quality improvement, measurement error, sample design, design of experiments, and the importance of collaboration and using mathematical methods to solve real problems. The theme of Hansen's research and books was identifying the essence of problems in censuses and sample survey design, and then finding practical and theoretically sound solutions to those problems.

The title of this article comes from the pet care guides that all of you have read at some time, for example "The Care and Feeding of Iguanas" and "The Care and Feeding of Puppies." The green iguana *(Iguana iguana)* needs natural sunlight, a balanced diet, and water to thrive. Puppies (*Canis lupus familiaris*) need a balanced diet, exercise, and socialization. The survey statistician (*Statisticus exemplus repræsentativus*) similarly needs proper care and feeding: a balanced diet, exercise, sunlight, and socialization. In this article, we examine these ingredients for survey statisticians. Of course, survey sampling is an interdisciplinary enterprise, and many areas of expertise are needed. My focus will be

[1] School of Mathematical and Statistical Sciences, Arizona State University, Tempe, AZ 85287-1804 U.S.A. Email: sharon.lohr@asu.edu
**Acknowledgments:** This work was partially supported by the U.S. National Science Foundation under grants SES-0604373 and DLS-0909630. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the National Science Foundation. I especially want to thank Burcu Eke for her assistance in collecting the sample of graduate statistics programs, and Snehalata Huzurbazar, Ellen Whitman, and Bryan Tronstad of the University of Wyoming for information and resources about Forest Hall.

on statistical and mathematical training at the university for statisticians who will be working on the surveys of the future. One ingredient is the sampling course, but other background is considered as well.

## 2. Balanced Diet: University Feeding

Before we start, let us remind ourselves of the essence of survey sampling. If you have a sample, it is easy to find the mean and other summary statistics for that sample. But how do you know whether that summary statistic has any validity? One of the fundamental problems of our discipline is how to generalize from what we observe to what we do not observe. In the eighteenth and nineteenth centuries, many philosophical societies debated the problem of induction. Among the philosophers and mathematicians who debated the question were Laplace, Kant, Peirce, Venn, and Quetelet. One subject for debate was how to estimate the probability that the sun will rise tomorrow; a common answer was to take the number of days the sun has risen, and divide that by the number of days the sun has risen plus one.

These philosophical discussions rarely affected official statistics, however, and recommendations by Bowley (1926) and others for random samples had little effect on most of the sampling being done in the early part of the twentieth century. At that time, although some people were starting to take samples instead of a census, these were often convenience or judgment samples. By necessity, the samplers had to *assume* that they had a representative sample. They thus were implicitly using a model where the nonsampled persons were assumed to be like the sampled persons. The samples were validated in some cases by comparison with the actual result, but no probabilistic framework was used to quantify uncertainty in the results. The famous example of the *Literary Digest Survey* (Literary Digest 1936) demonstrated the flaws arising from having no inferential framework. The editors claimed the survey was accurate because it had forecast every election correctly for 20 years. But in 1936, the *Literary Digest Survey* predicted that Landon would win with 55% of the votes when in fact Roosevelt won with 61%, and the 1936 election forecast was partly responsible for the magazine's demise.

Probability sampling provided a rigorous framework for making inferences from samples. It is easy now to forget what a revolutionary idea this was at the time. Neyman's (1934) groundbreaking paper argued against purposive selection of a sample, which was considered by many at the time to be by far the more logical way to select a representative sample (Anderson 1988, p. 183). After all, how could selection of a sample be relegated to chance? What if the sample happens to contain only the tallest people in the population? It seemed obvious that expert judgment would be needed to provide a representative sample. In probability sampling, inference is based on random variables $Z_i$, where $Z_i = 1$ if unit $i$ is included in the sample and $Z_i = 0$ if unit $i$ is not in the sample. The characteristics of interest $y_i$ are considered to be fixed; the $Z_i$'s are the only random variables present and the population total is estimated by $\sum Z_i y_i / P(Z_i = 1)$. You can see why this concept is challenging for students to understand: inference does not rely on intrinsic properties of the characteristic of interest, but rather on the probabilities that units are selected for the sample.

It is in this context that Hansen, Hurwitz, and Madow (1953) wrote the two volumes of *Sample Survey Methods and Theory*. The books are models of pedagogy and are driven by the practical problems of selecting and analyzing a sample. The theme of reducing bias and variance through use of a proper probability sampling design is emphasized in both volumes. Volume I begins with a discussion of biases and nonsampling errors in surveys to demonstrate the value of probability sampling for quantifying survey error. It then describes the basic probability sampling designs of simple random, stratified, cluster, and stratified multistage sampling. Ratio and regression estimation and double sampling are presented as methods for improving the efficiency of estimators. All of these topics are motivated by and integrated with survey problems. Volume I also introduces the random group method for estimating variances, which later led to replication methods such as jackknife and bootstrap.

Volume I concludes with three fascinating and groundbreaking case studies that illustrate how the principles of survey design are implemented in specific situations. The Sample Survey of Retail Stores, in Appendix A, describes an early use of multiple frame sampling. In the description of the Current Population Survey the authors review the practical reasons that motivated the design choice. Appendix C, on the Annual Survey of Manufactures, shows how optimal sampling fractions were derived for the survey and discusses use of ratio and regression estimation with the survey.

Volume II of Hansen, Hurwitz, and Madow (1953) then derives the mathematical results in Volume I. The last chapter of Volume II is on response errors in surveys, and is largely extracted from Hansen, Hurwitz, Marks, and Mauldin (1951), which, along with Deming (1944) and Mahalanobis (1946), initiated the study of measurement errors in surveys. Here again, the design of surveys is emphasized, particularly for assigning interviewer workloads so that contributions of interviewers to variance can be measured. For example, in the design of the Current Population Survey, the primary sampling unit (psu) size was chosen in part because it allowed one full-time supervisor over the interviewers in each psu; it was thought that having better supervision would reduce response errors due to interviewers.

Although Volume II does not present a full model-based approach to inference, models are used for the measurement errors. The model-based approach for inference in sample surveys was fleshed out starting in the 1960s; see Valliant, Dorfman, and Royall (2000) for a history and review. In a model-based approach, a stochastic model about the quantity $y_i$, considered in model-based inference to be a random variable, is used to predict the values of $y$ for units that are not in the sample. The population total is estimated by the sum of the $y_i$ for observations in the sample plus the sum of the predictions for observations not in the sample. A model-based approach is necessary for problems such as measurement error and nonresponse that are outside of the purview of a probability sampling design. In my view, a sampling student needs a thorough understanding of both approaches.

With that background in mind, let us look at the diet that students are getting in courses now. We took a simple random sample of 80 graduate programs in statistics or biostatistics in the United States, using the combined listings found at www.amstat.org as a sampling frame. The University of Michigan, University of Maryland, University of North Carolina, University of Nebraska-Lincoln, and Iowa State University were excluded from the frame because they have well-known programs in survey sampling. For each program, we ascertained whether the course catalog listed a graduate class in sampling (if it did not, we checked whether sampling had been offered as a topics class). We then determined whether the class had been offered in the last two years. Whenever possible, we examined the course web page and syllabus to determine the content of the class; otherwise, we used the official course description from the catalog. When more than one class was offered (which was rare), we examined all sampling classes.

We categorized a "basic" sampling class as one that covers simple random sampling, stratified, cluster, and multistage designs along with ratio and regression estimation – the basic class, in fact, covers only part of the material in Hansen, Hurwitz, and Madow (1953) Volume I. We also looked at other topics taught in one or more classes in the graduate program: replication variance estimation (which, remember, was in Volume I of Hansen, Hurwitz, and Madow 1953 through the random group method), regression and categorical data analysis with survey data, methods for dealing with nonresponse, spatial sampling, adaptive sampling, and model-based inference.

Very few sampling courses covered spatial or adaptive sampling, or explicitly mentioned model-based inference. The other topics from the 80 programs are displayed in Table 1. Of the 80 programs, 21 had no sampling class listed, and an additional nine that listed a sampling class in the catalog had not offered the class in the last two years. Twenty-two classes offered the "basic" syllabus of the standard probability sampling designs plus ratio and regression estimation (two of these, in fact, were half classes in sampling, combining sampling with categorical data analysis or design of experiments). The remaining 28 classes had at least one additional advanced topic, although we do not know the extent to which these topics were covered. Thus, of the 80 graduate programs, 30 offered no sampling class and an additional 22 offered only a subset of the material that was available in 1953.

How well does the sampling class prepare the student for doing sampling? Here is a list of some current topics in survey sampling discussed at the 2009 Joint Statistical Meetings:

- Weighting and weight smoothing or trimming
- Computer-intensive variance estimation
- Visualization
- Multiple-mode and multiple-frame surveys

*Table 1. Sampling topics taught in the 80 sampled graduate programs*

| Type of class | Number of programs |
|---|---|
| No sampling class listed for the program | 21 |
| Sampling class in catalog, not offered in last two years | 9 |
| Basic sampling class | 22 |
| Basic + replication variance estimation | 2 |
| Basic + regression or categorical data analysis | 7 |
| Basic + methods for dealing with nonresponse | 3 |
| Basic + replication variance estimation + nonresponse | 7 |
| Basic + variance estimation + regression or categorical + nonresponse | 9 |
| Total | 80 |

- Small area estimation and disease mapping
- Longitudinal surveys, and missing data in longitudinal surveys
- Coverage assessment and improvement
- Time series and spatial methods in surveys
- Nonparametric models for survey data
- Robust methods in survey sampling
- Sampling in rare populations
- Record linkage and efficient use of administrative data
- Empirical likelihood methods in surveys
- Confidentiality protection
- Nonresponse, imputation, and calibration

Of that list, possibly only the first, second, and last items are generally covered by the typical sampling class in a university. Of course, one class cannot include all the possible research topics, but it can give an idea of the real problems that exist. Students can discover these real problems themselves if they are given opportunities for exercise.

## 3. Exercise

The fledgling survey statistician needs exercise: I think it is essential for students to work with real data from a survey. Students in my sampling classes download survey data for a subject they are interested in from www.fedstats.gov. They read the codebook and technical documentation to learn about the survey design, and write a paper describing and critiquing the design and methods used to deal with nonsampling errors. The students are asked to graph the sample data and to provide graphs that incorporate the design information to estimate histograms and scatterplots of the finite population. They then perform univariate and multivariate analyses of variables they are interested in, using regression, logistic regression, or categorical data analysis methods for survey data. Sometimes the data sets include the stratification and clustering information so that Taylor series methods may be used for variance estimation. Some data sets include variables of replicate weights so that students analyzing those data need to understand the replication variance estimation method being used.

Many students choose to work on the National Health and Nutrition Examination Survey, often looking at the relationship between factors such as cholesterol levels and obesity. Other projects have included predicting the number of friends a student has using data from the National Longitudinal Study of Adolescent Health, studying the energy consumption of buildings with different energy-saving systems using the Commercial Buildings Energy Consumption Survey, and exploring the relationship between mathematics scores, gender, and calculator use with data from the Trends in Mathematics and Science Study (TIMSS). The TIMSS survey provides jackknife macros so the student must thoroughly understand variance estimation methods in order to properly analyze the data.

If resources permit, it is even better to have students work on all the steps of designing and analyzing a survey. A survey research center at the university is helpful, but not necessary. Students can design and select samples from Internet data, for example, sampling books from the listings at amazon.com, so that they obtain practice in carrying out a survey. By sampling from the Internet, they encounter many of the problems from real surveys but without the expense and difficulty of sampling persons. When sampling books from amazon.com, for example, some values are missing and there is measurement error. Even the simple random sample displayed in Table 1 had a number of challenges for coding. Another option is to treat a large data set as a population and take samples using different designs from that population.

By having these projects, rather than exercises that are carefully crafted to emphasize one concept at a time without additional complications, students learn about survey sampling as currently practiced. They also start thinking about what new methods are needed, and what is not working with the design. An advantage of having students work on large surveys in the class is that they discover for themselves the practical problems involved in current survey methods; this suggests research problems to students.

## 4. Sunlight

We want to prepare students for the research problems of the present and future in sampling. In Section 2, I mentioned some of the current challenges in sampling. Predicting the future is never easy, but we can look at history for a guide to trends in the development of the subject.

Throughout the history of survey sampling, innovations have been spurred by technological advances and societal changes. In the 1930s and 1940s, two of the spurs for the explosion in survey research were the economic depression and the war, with the attendant needs for more and better statistics. Estimating unemployment and its effects was imperative. World War II increased receptivity and demand for accurate statistics, and the U.S. Census Bureau became involved in the war efforts.

Hansen was also concerned about the errors in surveys. According to Anderson (1988), when Hansen came to the U.S. Census Bureau, errors in the census were viewed as unfortunate but unavoidable. Hansen, Hurwitz, Madow, and Deming sought to change this, and developed mathematical theory for calculating sampling errors. In Anderson's words (p. 194), "If certain census methods were prone to error, then better census methods would have to be devised."

The 1960s were also a time of great societal and technological change. Telephone surveys were increasingly used, which led to the development of new designs such as the Mitofsky-Waksberg cluster sampling design. More attention was also paid to measurement errors, especially in the classic paper by Hansen, Hurwitz, and Bershad (1961). The 1960s and 1970s also saw the development of model-based inference, which viewed survey inference as a problem of prediction of unobserved values in the population.

Better computing has also spurred many of the innovations in sampling. Hansen was a pioneer in recognizing the potential of the new electronic computing machines and was instrumental in procuring the UNIVAC computer for the U.S. Census Bureau in 1951 (see Hansen 1987). The new computing made it possible to improve the collection and analysis of survey data, and to use more advanced probability sampling designs. Replication variance estimation methods, of course, were only made possible by the possibility of computing them, and the area saw an explosion of research in the 1980s. In addition, better computation as well as society's need for more information from the surveys led to the development of theory and methods for carrying out regression and categorical data analyses on data from complex surveys, as well as small area estimation methods.

The 2000s have, of course, seen a tremendous increase in computer power. They have also seen the development of the Internet and mobile devices as a primary means of communication. Massive amounts of data are available, but the data are often of uncertain quality. Data are often sampled from various sources and are collected through different modes. Multiple sources, multiple phases, and multiple modes all increase the challenges for producing estimates.

It certainly seems as though the Internet must be part of the future for sample collection. The problem is that probability sampling designs that work well for other modes often do not work well for Internet samples. Although data collection and processing is inexpensive, the nonsampling errors are a serious problem. Sampling frame construction, coverage assessment, and nonresponse are among the challenges that must be met for an Internet survey to have high quality. Measurement error in Internet surveys is another subject needing a great deal more research. How to use probability sampling in an Internet age represents a challenge for statistical theorists: as Hansen, Hurwitz, and Madow (1953) express it, there is "[o]pportunity for ingenuity in sample design" (Volume I, p. 456).

Remember, in the 1920s there was no firm mathematical foundation for survey inference, and polls claimed accuracy because of a recent track record in correctly predicting election outcomes. Now, some polls have a response rate of about 15%. Some organizations use volunteer (or paid) online panels for polling. Many of these, then, are a return to convenience samples. Yet they report a margin of error. The margin of error presupposes a basis for inference, but what is that basis? No probability sample was taken, so probability sampling cannot be used as a basis for inference. Yet there is no evidence of careful modeling that would allow a model-based perspective to be adopted; with a sample of volunteers, how could one have any confidence that a model holding for the sample holds for the population?

This is a disturbing trend, since the public judges all of survey research by polls. Recently, various organizations have begun rating polls by their accuracy: how close was the poll to the election? Some more careful polls compare respondents to a subset of the nonrespondents. However, I think the polls need to be evaluated by methodology, not past accuracy. Polls without an inferential framework should not report a margin of error. Indeed, recent research indicates that the claims of accuracy by organizations that use nonprobability sampling may be overstated. Yeager et al. (2009) compared probability and nonprobability samples and found that the probability samples had uniformly better accuracy. They also found that poststratification did not necessarily help compensate for the selection bias of the nonprobability samples.

It is always dangerous to try to predict the future, but let us look at some other current trends that may affect survey research. As mentioned before, coverage, nonresponse, and measurement error continue to be challenges. At the same time, more and more information is available about individuals and regions that can be used as auxiliary information. The data are sometimes networked; in other circumstances data are available from multiple sources and must be linked. Given that much of the theory of survey sampling has paralleled the development of computational capacity, it seems reasonable to assume that this will continue.

There is a danger that the ready availability of auxiliary data may lead some people to think carefully designed surveys are unnecessary. Wilkinson (2008) mentions a side-effect of user-friendly software for structural equation modeling in the social sciences: the number of correlational studies has increased at the expense of the number of designed experiments. Haphazard data, however, have no controls on data quality; as with online panels, heroic assumptions must be made in order to make inferences about the population. There may also be little control over measurement error and other nonsampling errors. Thus, we need to continue to advocate carefully designed and properly analyzed surveys, and be vigilant in promoting high-quality survey operations. As Hansen, Hurwitz, and Madow (1953, Volume I, p. 73) state: “Since we know of no statistical theory for measuring the reliability of sample results by purposive or other nonrandom sampling methods, such methods are automatically excluded if the criteria of good sample design described in Chapter 1 and adopted in this book are followed.” When one uses nonprobability methods, one should keep in mind that one “is obtaining results whose accuracy must be based on assumptions and judgments that cannot be measured objectively” (p. 73).

Surveys collected by federal statistical systems form some of the highest quality data available anywhere. If the past is any guideline, the demands and uses for those data will only increase. I cannot, of course, predict anticipated uses for survey data, but I think it likely that in the near future, data may be used for detecting anomalies such as disease clusters and that new methods for using partial information will be needed.

The training of survey statisticians, then, should not just consist of teaching students dry formulas for estimating totals and variances. We want to follow the example of Hansen, Hurwitz, and Madow (1953) of using theory to solve real problems of how to select and analyze samples, but using the methods available to us in the 21st century. This will give students a framework of exploration and problem-solving that they can use to address the sampling needs of the future.

## 5. Socialization and Training for the Future

Much of mathematical statistics training starts with the assumption that data are independent and identically distributed (iid). This is a reasonable place to start, since the iid assumption makes theory tractable for students and beginning with iid data mirrors the historical development of statistics. But many of the open problems in statistics today involve clustered data: spatial problems, repeated measures data, network data. A background in sampling gives all statisticians an advantage for working with dependent data. I argue that better socialization between disciplines benefits researchers in all areas.

One benefit of improved training in survey sampling will be increased use of sampling methods in other fields. We are surrounded by huge data sets with relationships waiting to be discovered. Friedman (1997, p. 7) suggested that “Sampling methodology, which has a long tradition in statistics, can profitably be used to improve accuracy [in data mining] while mitigating computational requirements.” In his Introductory Overview Lecture at the 2009 Joint Statistical Meetings, D. J. Hand mentioned modeling issues in data mining. He said that many people recommend just taking a sample, and analyzing that to discover relationships, but said that simple random sampling may be difficult or impossible. Survey statisticians, however, have many more tools than just simple random sampling, and this seems to be an area where we could make a profound contribution. Even though data collection may be dynamic, and data may be dispersed on servers around the world, survey statisticians know how to address these types of problems. Survey sampling provides a valuable background for working in the area of data fusion; with its focus on inference to finite populations, it provides a different perspective for work in integrating data from different sources.

Other problems, such as comparing medical treatments, can also benefit from probability samples. Small area estimation methods can be applied in many areas, and one of those areas is examining effectiveness of medical treatments for small demographic subgroups: the sample size may not be sufficient for direct inference in a demographic subgroup, but we can "borrow strength" from other subgroups.

What I am saying about how the entire discipline of statistics can benefit from more connections with survey sampling is not new. Tukey (1962), in his classic article "The Future of Data Analysis," lists survey sampling as one of the necessary tools for doing statistical research:

> It is, incidentally, both surprising and unfortunate that those concerned with statistical theory and statistical mathematics have had so little contact with the recent developments of sophisticated procedures of empirical sampling. The basic techniques and insights are fully interchangeable with those of survey sampling, the only difference being that many more "handles" are easily available for treating a problem of statistical theory than are generally available for treating a problem about a human population or about an aggregation of business establishments (p. 61).

We need, therefore, to train future survey statisticians for a wide variety of potential applications that are, at present, unknowable. Efron (2007) argues that the discipline of statistics "is in a period of rapid expansion and change. During such times, it pays to concentrate on basics and not tie oneself too closely to any one technology or analysis fad." Binder (2000), discussing training for survey methodologists at Statistics Canada, emphasizes that although statisticians at the agency work in applied statistics, the most important part of their training is the foundation in statistical theory. He also encourages exposure to practical problems and writing reports.

With these wise words in mind, I think we need core training for *all* statistics students that emphasizes statistical thinking and gives flexibility for what may be required in the future. We have many challenges in survey sampling that desperately need new statistical theory. And since we do not know what new challenges will arise, we need students with a variety of backgrounds and training. There are many parallels between our current situation and the situation Morris Hansen faced in the 1930s: there is, of course, the economic crisis and the attendant need for more and better statistics. But there is also a parallel that, in many applications, the survey theory is struggling to catch up to the needs. We know that new theory is needed, but do not know what form it will take; therefore, we need researchers who can use methods from many areas to explore new ideas. There could well be a new paradigm for survey inference around the corner. After all, who in 1920 foresaw the changes that would come as a result of probability sampling?

Figure 1 sketches my view of a common graduate statistics curriculum. It requires a number of courses in the mathematical theory of statistics. The mathematical results are

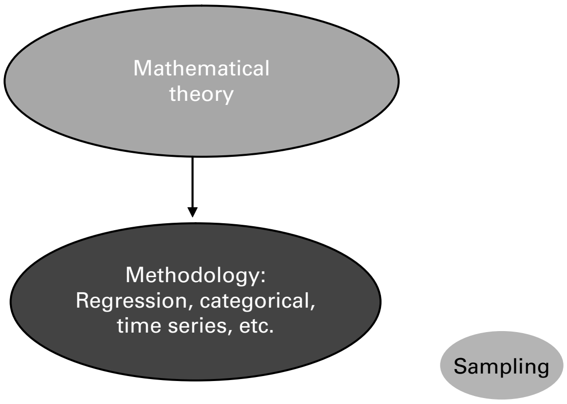


*Fig. 1. A typical graduate statistics curriculum. The mathematical theory is the central core, and students are taught how statistical methodology derives from the mathematics. The sampling class, if it exists, is often not integrated with the other classes*

then applied to statistics methodology classes such as regression, categorical data analysis, time series, and so on. The flow is often one way only, as depicted by the arrow in Figure 1: ideas from mathematics are applied in the methods courses, but students often do not see the other direction of flow in their classes. It is the reverse direction, however, that leads to new statistical theory that is relevant to real problems. Students working on data sets and design problems encounter a new situation and set about using their knowledge to devise new statistical methodology.

In many graduate programs, when there is a sampling course, it stands alone: while other courses are integrated with the other theoretical and applied curriculum, the sampling class appears isolated and no other class requires the sampling class as prerequisite. Even when graduate statistics students are required to take a sampling class, it is easy for them to get the impression that this is a quirky and outdated subject completely disconnected from the rest of statistics.

Today's and future survey samplers still need a solid foundation in the mathematical theory for statistics; in statistics methodology such as regression, categorical data analysis, and time series; and in probability and model-based sampling. But the curriculum can be updated so that the emphasis is on solving problems using statistical thinking. Figure 2 shows my view of a graduate curriculum in statistics. Statistical thinking and data structure are in the center of this curriculum, and all of the theory and methodology courses reflect this core. The core naturally includes socialization with people in other fields and developing statistical methods to solve real problems. Hansen, in his May 1986 interview with Ingram Olkin (1987), was asked to give advice for the new statistician. Hansen replied:

> I feel that it's just wonderful what you're doing in education – interacting in a subject area with statistics. Some of the people I've seen coming out in mathematical statistics with Ph.D.s are more interested in writing papers than in solving applied problems. If they're interested in applied work and have a reasonably good statistical background, they can learn to work effectively in almost any field. For example, Bill Cochran moved from experimental design to sampling to biostatistics, back into a more general sphere.

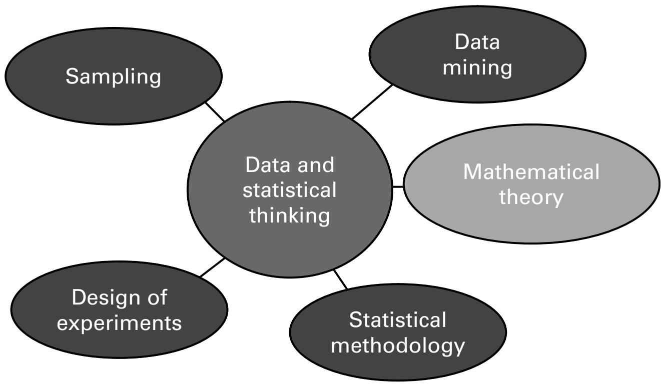


*Fig. 2. A data-centric statistics curriculum. The aspects of statistics dealing with data collection – sampling and design of experiments – are given much more emphasis*

> I am sure that advancing theory as a basic research activity, and publishing papers, is exceedingly important, and has made major contributions to applied work. But there is also a need for people who have learned to want to solve applied problems and not just write papers (p. 177).

The curriculum in Figure 2 places great emphasis on data collection methods through sampling and design of experiments. This too reflects Hansen's view of the importance of designed experiments for improving survey quality. As early as 1950, Hansen was using randomized experiments at the U.S. Census Bureau to measure within- and between-enumerator components of variance (Olkin 1987). Studies of measurement error and quality improvement rely heavily on designed experiments (Hansen and Waksberg 1970). Data mining methods are included as well: we are surrounded by massive data sets and we need to both develop methods for sampling those data and develop methods for using that information as auxiliary information in surveys.

Better integration of theory and methodology benefits both the sampling class and the other classes in the curriculum. Most students take classes in asymptotic theory at some point in graduate school. This is an ideal place to talk about probability sampling inference and how it differs from the "usual" theory; Lehmann (1999), for example, has both design-based and model-based central limit theorems. Missing data arise in many settings, and complex surveys can be used as examples in a missing data class.

Nolan and Temple Lang (2009) argue that computing needs to play a much more prominent role in statistics education, and I wholeheartedly agree. This is especially true in survey sampling, where the data sets are large and, increasingly, replication variance estimation methods are used. Nolan and Temple Lang (2009) advocate having computational statistics classes in the graduate curriculum, and survey research should be an integral part of such classes. The sampling class can reinforce the rest of a modern curriculum by being data-centric and teaching students how to develop statistical methods to solve problems. Increased use of computing in sampling classes allows more class time to be spent on topics outside of the basic syllabus such as nonresponse or sampling for rare events.

We also need to recognize that we cannot expect one student to know or do everything. Genetic diversity is tied to species survival, so we should encourage students to follow their interests into learning about other areas. I think that all statistics students should learn about sampling in graduate school. Some students should also study machine learning, graph and social network theory, bioinformatics, and many other topics. Better socialization within the statistics curriculum, and better socialization with other disciplines, will give survey statisticians a flexible background that will allow them to adapt and thrive in the environments of the future.

## 6. A Thriving Species

But perhaps we should worry that the species *Statisticus exemplus repræsentativus* may be endangered. I listened to the Senate Confirmation Hearing of Robert Groves from May 15 on C-SPAN (2009). Senator Akaka from Hawaii stated that: "The federal government is facing major human capital challenges . . . You estimate that 45% of current Census employees will be eligible to retire next year . . . . What steps will you take to ensure that the Bureau is able to recruit and train workers with the technical background needed to replace these retirees?" Groves replied: "I am terribly worried about this problem. It is not a problem only of the U.S. Census Bureau but of the entire federal statistical system . . . . Secondly, the number of programs in the country training people that have the requisite skills for the U.S. Census Bureau is way below the need."

I think part of the reason for the shortage of programs training people with these requisite skills is the relative isolation of our discipline. Figure 3 displays the geographic distribution of members of the Survey Research Methods Section (SRMS) of the American Statistical Association. This is not a perfect measure, but I am using it as a proxy for interest in statistical survey research. About 34% of the 1,241 members in the United States list an address in the Washington, D.C. area. Within states, too, the membership is highly clustered. Just as many graduate programs in statistics have no class in sampling, many graduate programs have no SRMS members among the faculty. This map is for the United States, but I suspect a similar clustering of survey statisticians occurs worldwide.

Which brings us back to Morris Hansen. The federal government faced a similar problem in the 1930s. According to Eckler (1972, p. 19): "In 1933, the Bureau had only three Ph.D.s and only one professional man under forty-five years of age." Anderson (1988, p. 174) reports that the 646 permanent employees of the U.S. Census Bureau in 1934 had average age 48.9 years. And, "bureau personnel were completely ignorant of the developments in probability theory that had revolutionized the science of statistics since the turn of the century. Bureau officials did not understand or use sampling theory, correlation coefficients, or calculations of probable error" (p. 175).

Morris Hansen graduated from the University of Wyoming with a degree in accounting in 1934. (The University of Wyoming, incidentally, has produced an unusual number of pioneering survey statisticians: W. Edwards Deming graduated from University of Wyoming with a BSc in electrical engineering in 1921; Edward C. Bryant received a master's degree from University of Wyoming in 1940; see Amstat News 2008). How did Hansen, whom Fellegi (1991) has called "the most important person in the development of survey methodology and applications," end up in survey statistics? Here are Hansen's words, from his interview with Ingram Olkin:

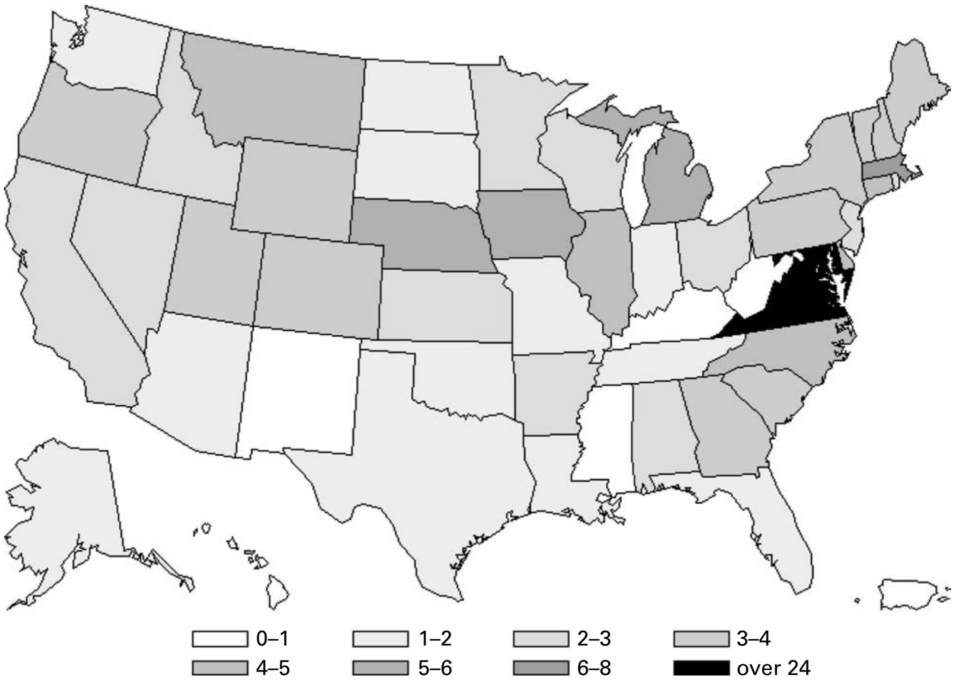


*Fig. 3. Geographic distribution of American Statistical Association Survey Research Methods Section members, showing the number of members per 1,000,000 population in each state*

> Well, in Wyoming I didn't know what I wanted to do and finally decided to take accounting after one false start. And in accounting I was exposed to courses in economic statistics by a professor in the Commerce Department. He was a really fascinating teacher and got me interested in statistics. When I finished those courses, I thought I knew something about statistics and learned later that was a misconception. But I knew a little and decided that I would like to go into statistics (Olkin 1987, p. 162).

Forest Hall, the abovementioned fascinating teacher, seems to have been an extraordinary individual. Some details about his life and work are given by Clough (1965) and The Laramie Daily Boomerang (1984). After serving in France during World War I and receiving a bachelor's degree in civil engineering from the University of Nebraska and a master's degree in commerce from the University of Denver, Hall was hired by University of Wyoming as an assistant professor in the Commerce Department (one of five full-time faculty members in the department) in 1927. During the depression, he served as a regional director for the U.S. Department of Labor. In 1935–1936, Hall took a leave of absence from the University of Wyoming to be supervisor for the Rocky Mountain region of the Study of Consumer Purchases, an early survey collected by the Department of Labor and other agencies (see Moen and Gratton 1999 for a description of the survey). He was also instrumental in establishing the "IBM bureau" in the summer of 1950 for data processing at the University of Wyoming – this was part of the division of studies and statistics, which he directed.

There you have it: Hansen became a survey statistician not because it was his lifetime dream to become a statistician, not because he was convinced by statistics about the lucrative career belonging to statisticians, but because he encountered a great teacher who was involved in sampling problems. Many of Hansen's interests in sampling – the importance of good design, the practical aspects of taking a sample, and the importance of having good computing – are reflected in Forest Hall's career.

We cannot train survey statisticians unless we attract them in the first place. You all know the saying that the word *data* is not the plural form of the word *anecdote*. But recruiting students to be survey statisticians relies primarily on anecdotes rather than on data. Students go into survey statistics because of personal connections, from being involved in and learning about activities that allow them to see the importance and excitement of the subject, and from great teaching.

We need to encourage great teaching of sampling at all levels: in grade school and high school (Statistics Canada has wonderful activities for having high school students participate in a survey, then compare their results to those of other students in Canada and around the world), in introductory statistics classes, in undergraduate programs, and especially in the graduate curriculum for statisticians. A number of universities are doing a wonderful job at recruiting and training survey samplers. But, in too many statistics programs, students can graduate without ever being exposed to a problem in survey sampling.

Survey sampling, however, is an integral part of a data-centered curriculum for students, one that trains not only future survey researchers but all statistics students in the art and science of data collection. Brian Joiner has used the term *numerical detectives* to describe statisticians, and I think this exactly captures what our discipline is about. Survey statisticians are the detectives who determine how reliable evidence about unemployment, health, the environment, crime, number of caribou – evidence that will stand up in court – will be gathered. The challenges of collecting reliable evidence change with time and technology, but we can safely argue that the need for persons who have the flexibility and background to create and welcome new inferential frameworks can only be expected to increase.

## 7. References

Anderson, M.J. (1988). The American Census: A Social History. New Haven, CT: Yale University Press.

Amstat News (2008). Obituary: Edward C. Bryant. Amstat News, 369 (March), 40.

Binder, D.A. (2000). Preparing for a Career as a Survey Statistician (at Statistics Canada). Proceedings of the Survey Methods Section, Statistical Society of Canada, 33–34.

Bowley, A.L. (1926). Measurement of the Precision Attained in Sampling. Proceedings of the International Statistical Institute, 12, 6–62.

C-SPAN (2009). Confirmation Hearing for Census Director Confirmation, Senate Governmental Affairs Committee, May 15, 2009, Product ID 286367-1, www.c-spanarchives.org/library/index.php?mainpage=productvideoinfo&productsid=286367-1

Clough, W.O. (1965). The History of the University of Wyoming, 1887–1964. Laramie: University of Wyoming.

Deming, W.E. (1944). On Errors in Surveys. American Sociological Review, 9, 359–369.

Eckler, A.R. (1972). The Bureau of the Census. New York: Praeger.

Efron, B. (2007). The Future of Statistics. Amstat News, (September), 47–50.

Fellegi, I. (1991). Obituary: Morris H. Hansen, 1910–90. Journal of the Royal Statistical Society, Series A, 154, 352–353.

Friedman, J. (1997). Data Mining and Statistics: What's the Connection? In Computer Science and Statistics: Proceedings of the 29th Symposium on the Interface, 1–7.

Hansen, M.H. (1987). Some History and Reminiscences on Survey Sampling. Statistical Science, 2, 180–190.

Hansen, M.H., Hurwitz, W.N., and Bershad, M. (1961). Measurement Errors in Censuses and Surveys. Bulletin of the International Statistical Institute, 38, 359–374.

Hansen, M.H., Hurwitz, W.N., and Madow, W.G. (1953). Sample Survey Methods and Theory, Vols. 1 and 2. New York: Wiley.

Hansen, M.H., Hurwitz, W.N., Marks, E.S., and Mauldin, W.P. (1951). Response Errors in Surveys. Journal of the American Statistical Association, 46, 147–190.

Hansen, M.H. and Waksberg, J. (1970). Research on Non-sampling Errors in Censuses and Surveys. Review of the International Statistical Institute, 38, 317–332.

Laramie Daily Boomerang (1984). UW Professor Emeritus Forest R. Hall dies. The Laramie Daily Boomerang, 104 (March 20), 10.

Lehmann, E.L. (1999). Elements of Large-Sample Theory. New York: Springer.

Literary Digest (1936). Landon, 1,293,669: Roosevelt, 972,897. The Literary Digest, 122 (October 31), 5–6.

Mahalanobis, P.C. (1946). Recent Experiments in Statistical Sampling in the Indian Statistical Institute. Journal of the Royal Statistical Society, 109, 325–370.

Moen, J.R. and Gratton, B. (1999). Tracking the Majority: Households, Older Workers, and Retirement During the Great Depression. Journal of Gerontology, 55B, S28–S32.

Neyman, J. (1934). On the Two Different Aspects of the Representative Method: The Method of Stratified Sampling and the Method of Purposive Selection. Journal of the Royal Statistical Society, 97, 558–625.

Nolan, D. and Temple Lang, D. (2009). Approaches to Broadening the Statistics Curricula. Quality Research in Literacy and Science Education, M.C. Shelley, II,, L.D. Yore, and B. Hand (eds). New York: Springer, 357–381.

Olkin, I. (1987). A Conversation With Morris Hansen. Statistical Science, 2, 162–179.

Tukey, J.W. (1962). The Future of Data Analysis. Annals of Mathematical Statistics, 33, 1–67.

Valliant, R., Dorfman, A.H., and Royall, R.M. (2000). Finite Population Sampling and Inference: A Prediction Approach. New York: Wiley.

Wilkinson, L. (2008). The Future of Statistical Computing (with Discussion). Technometrics, 50, 418–447.

Yeager, D.S., Krosnick, J.A., Chang, L., Javitz, H.S., Levendusky, M.S., Simpser, A., and Wang, R. (2009). Comparing the Accuracy of RDD Telephone Surveys and Internet Surveys Conducted with Probability and Non-probability Samples. Technical report, comm.stanford.edu/faculty/krosnick/Mode 04.pdf

Received October 2009