Library & Information Science Research 33 (2011) 92–95

Contents lists available at ScienceDirect

Library & Information Science Research

Statistical correctness ☆

Ray Lyons

4195 West 59th Street, Cleveland, OH 44144, USA

ARTICLE INFO

Available online 18 December 2010

ABSTRACT

This paper calls for improved survey research methods both in formal research projects and in everyday use of survey questionnaires in libraries. The author proposes that probability sampling be recognized as the best practice in library and information science. The problem of selection bias arising from the use of nonprobability sampling, and convenience sampling in particular, is explored. The paper presents examples of unfounded research conclusions appearing in studies using this sampling approach. It also espouses library research and assessment practices that aspire to higher standards of soundness, accuracy, validity, and impartiality.

1. Introduction

When our information need happens to be evidence to serve as a foundation for professional practice, or data about the quality, effectiveness, and worth of library services, we in the library and information profession sometimes prefer convenience and expedience over accuracy and thoroughness. Like the most impatient of information seekers, we ignore the fact that inadequate information gathering techniques will lead us quite expediently to the *wrong answers*. Neither do our national and international library organizations set consistently good examples in this regard. Too often they employ deficient research methods or promote unjustifiable interpretations of data they have collected. Unfortunately, these practices are emulated by libraries which are unaware of the methods' shortcomings.

2. Convenience sampling

One substandard practice is especially problematic due to its popularity among library organizations: the use of convenience sampling in survey research. This is an overstatement, of course, but my advice to a library organization planning to conduct a survey using convenience sampling is that they consider fabricating the data instead. While their fabrications will almost always be wrong, the organization will not be so careless as to give the data more credence than they deserve. On the other hand, surveys that use convenience sampling are (except for certain special cases) also quite likely to be wrong for reasons to be explained. Yet the tendency is to consider the results to be accurate and reliable. Driven by data like these, library organizations are liable to veer away from their desired destinations.

Most readers will be aware that survey samples are used to gather data describing a given group we are interested in (population) without having to conduct a complete census of this population. Convenience sampling belongs to a family of methods that statisticians call nonprobability sampling.[1] The main failing of convenience (and nonprobability) sampling is its inability to provide a complete and balanced picture of the larger population of interest.

For readers new to the idea of convenience sampling I offer the following example: Suppose a library consortium wants to measure user opinions about reference services provided by member libraries. For this purpose a consortium researcher decides to visit the library that happens to be located downstairs from the consortium's office. She proceeds to watch for and then interview patrons who have just completed an interview with reference librarians currently on duty. She does this from 10 a.m. to 11 a.m. every day for 1 week since her schedule is free during those hours of the week.

Assuming her interviewing technique to be excellent, the researcher will learn about users in this library, in the town where the consortium offices are located, who typically visit the library at this time of day and this time of year, and who are inclined to talk to her. But these interviewees will not speak for the complete population of users of reference services provided by the libraries—for example, users visiting this library in the evening or on weekends, visiting other consortium libraries, or using telephone or electronic chat reference

☆ The following article is a departure from the norm for Library and Information Science Research, and we welcome reader response and comments (address to either of the editors).

E-mail address: raylyons@gmail.com.

[1] Statisticians consider chance as a fair and unbiased method for selecting some members from a population and excluding others. Sample selection methods that conform with this principle are called probability sampling. Sample selection methods that do not are called nonprobability sampling.

0740-8188/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.lisr.2010.07.016

services, whose reference questions might be extraordinarily complex, and so on.

The best practice in survey sampling is probability sampling. This family of sampling methods consistently takes the entire population of interest into account. As a result, these methods broaden the representativeness of the data collected and add to its meaningfulness. Looked at another way, these methods minimize sources of bias in the data collected. A probability sample for a survey of the consortium's patrons concerning reference services would include a complete range of libraries and towns, a range of users regardless of their frequency of or schedules for visiting the library or asking reference questions, or the complexity of their questions, users of telephone, e-mail, and electronic chat reference services, users receiving services from a variety of reference staff, during various hours of operation, and so forth.

Because probability sampling avoids pitfalls that introduce bias into survey research, it produces more comprehensive and accurate information. Obviously, gathering high quality information requires an investment of time, effort, and money, as does, for instance, assuring the accuracy and completeness of our institutions' financial accounts. Consideration of the procedures, effort, and expense involved in probability sampling requires a more in-depth exploration of this family of sampling methods than can be addressed here.

3. Selection bias

Statisticians refer to the problem of sample unrepresentativeness as selection bias, or more generally as selection anomalies. The point is that the initial definition of the pool of subjects to be studied leads researchers down a pathway to wrong answers. Statistician Howard Wainer has a delightful chapter about this topic in his book (Wainer, 2005). He recounts a 19th century Swiss study of the longevity of various professions as defined by average age of death for practitioners of each based on official records (Wainer, 2005, pp. 143–144). The study's results included the peculiar finding that the average age of death for the profession "student" was found to be 20.7 years. The Swiss researchers did not realize that the mistake leading to this wrong answer also affected all of the average ages reported in the study. The mistake is using an insufficient and unrepresentative pool of subjects. Wainer (2005) also cites more contemporary examples of oversights of this type, including a 1991 American Podiatry Association study claiming that 88% of all women wear shoes that are too small (p. 144).

Using convenience sampling leaves us in the dark about the accuracy and relevance of survey results. Consider a longitudinal study of school libraries conducted by the American Association of School Libraries (AASL) that used convenience sampling (AASL, 2007, 2008, 2009). In each study year respondents were self-selected, meaning that any school libraries volunteering to respond would be included. Just as podiatry patients are likely to report foot pain, so will self-selected respondents report signs and symptoms that nonrespondents may find unfamiliar or irrelevant. As with other types of convenience sampling, self-selected samples produce data that are almost always biased. Without evidence that the respondents represent the larger population fully and fairly, we must conclude that they represent only themselves. Statisticians describe this situation by saying that the survey findings are not generalizable to the entire population, or that there is no statistical basis for drawing inferences from the sample data to the entire population.

Surely the AASL researchers understood that none of the survey results was generalizable to the school libraries nationally. However, their reports neglected to clearly explain this. Instead, the first two reports stated only that respondent self-selection prevented "generation" of "national totals" from the data (AASL, 2007, p. 1, 2008, p. 2). To illustrate this problem, the authors noted that two thirds of the U.S. states from which respondents came accounted for three fourths of all survey respondents[2] (AASL, 2007, p. 1). This, along with the lack of further elaboration, left the impression that this geographical imbalance limited some, but perhaps not all, of the survey's generalizability.

Almost universally, reports from surveys that use convenience sampling contain qualifying phrases like "X% of survey respondents reported ..." or "Y% of librarians surveyed said ...". The problem is that as these reports proceed the phrases and their implications recede into the background where they are eventually forgotten. As we might expect, the 2007 and 2008 AASL reports made liberal use of these phrases. However, by the 2009 report they disappeared altogether. Worse, nowhere in that report do we read that convenience sampling was used or that the study data were subject to bias (AASL, 2009). Rather, the report consists primarily of unfounded statements about school libraries nationwide (since the survey data are not generalizable).

The AASL reports would be so much more valuable if they clearly disclosed the strengths and weaknesses of the data and allowed readers to decide for themselves how relevant the findings are. As it is, their latest report portrays the longitudinal data as reliable and robust. That edition happens to include margin of error figures for all three study years (AASL, 2009, p. 3). A margin of error is an estimate of how precise survey data are likely to be based on certain statistical assumptions[3] (Scheuren, 2004, p. 63). Specific statistical rules apply to calculating margins of error, including the requirement that the data be selected by means of probability sampling (Scheuren, 2004, p. 66). We have no dependable ways for estimating the precision of data that come from convenience samples. This is the message research and evaluation studies using this sampling method ought to communicate.

4. The problem of nondata

The ease of creating online questionnaires has made convenience sampling even more convenient, while also increasing the temptation to conduct surveys without sufficient forethought and planning.[4] For instance, last winter in a Midwestern U.S. state, library organizations hastily promoted an online questionnaire in preparation for an application for federal stimulus money for broadband Internet expansion. Public libraries enthusiastically posted a link to the questionnaire on their websites, along with an announcement of a contest to "Win a Netbook Computer!" We might wonder, however, which library users this survey will attract. Who will be over- and underrepresented? How likely are users with limited computer skills to respond? Those with excellent skills? Those with accessibility difficulties? How many rural residents with dial-up Internet connections will take time to respond? And so on.

Wainer (2005) applies the term "nondata" to data collected using insufficient methods like convenience sampling (p. 57). Because we know so little about who does and does not respond to these surveys, the data are nearly worthless. In the worst cases, they are in the same league as results from online polls in which Internet users express their opinions (as often and as erratically as they wish) concerning the latest celebrity gossip, political scandal, and such. While this

[2] In a probability sample each U.S. state would be represented in proportion to the number of school libraries within its boundaries. An imbalance in responses among states is not necessarily grounds for disqualifying the sample from representing school libraries nationally. The study's use of convenience sampling is the basis for this disqualification.

[3] Inaccuracy in survey results due to sampling imprecision (or variability) is called sampling error. Inaccuracy from other sources like confusing or biased survey or interview questions, respondent duplicity, data tallying errors, and so forth is called nonsampling error. Survey margins of error do not detect or estimate nonsampling error (Scheuren, 2004, pp. 63–66). See also Thompson (2006, pp. 200–206) and Hays (1973, pp. 375–380).

[4] For a primer on best practices in planning and implementing surveys see Scheuren (2004).