

From: **Journal of Education for Librarianship. 1978 Summer; 19(1):3-20**

Teaching Descriptive and Inferential Statistics in Library Schools

Ronald E. Wyllys

Associate Professor, University of Texas Graduate School of Library Science, Austin TX 78712

Abstract

The goals of the paper are: (1) to suggest that library and information science already depend heavily on statistical techniques; (2) to present examples of the need for greater knowledge of statistics, especially inferential statistics, among librarians and information scientists; and (3) to argue that library school students can be taught inferential statistics. In support of (1) and (2) a study of the 1,157 articles published in 1975 in 36 library and information science journals showed that about 45 per cent of the articles made some use of statistics and that about 3 per cent of the articles, concentrated in the leading journals, used inferential statistics. Not only does the latter percentage indicate that authors missed opportunities to make more penetrating analyses of many sets of data, but also it is related to mistakes in the conclusions that some authors draw from their descriptively analyzed data. In support of (3), a description is given of the research methods course at the author's institution, in which students learn inferential statistics with computer assistance.

Should library schools offer courses in quantitative methods and other methods of research? How extensive should the treatment of statistics in such courses be? The object of this paper is to argue that to the first question the answer should be, "Yes," and to the second, "Not only descriptive but also inferential statistics should be taught."

Numerous articles in recent years have dealt with the importance of gathering and reporting statistics on various aspects of libraries, as well as with the need to improve the accuracy of such statistics. The 1970 book¹ edited by D. C. Palmer furnishes an excellent introduction to these problems, and a paper² by T. A. Childers in 1975 summarizes later work. Other recent papers by M.E. Anders,³ R. W. Burns⁴ and M. A. Drake⁵ have treated particular uses of library statistics.

Two especially frank articles, by S. Leech⁶ and W. H. Williams,⁷ spell out the importance of library statistics for a highly practical purpose: justifying the library's existence. Williams sums up the matter with the words, "the financial success of any library, which ultimately is the true test of services, results largely from your ability to prove statistically your worth to the community" that uses, or could use, the library.

The emphasis in most of the works just cited is on the collecting of library statistics and on the problems involved in doing so. It is to be hoped that the eventual success of such efforts as the LIBGIS (Library General Information Survey) project⁸ will lessen the need, in the library and information science literature of the future, for papers preaching the importance of collecting such statistics. It is time for the profession to place more emphasis on utilizing the statistics it collects.

In her paper Leech treats, succinctly and wittily, reasons for not merely collecting but also analyzing library statistics. She points out that in extracting useful information from such statistics librarians are "limited only by [the] time and imagination" they bring to bear on the task. "You will have information that can help you tell your own departmental personnel whether they are up to snuff, you can help them plan ways in which they might improve their performance, [and] you can analyze department deficiencies in work flow"⁹

In a similar vein, Anders writes: "Generally, as Purdy observed, librarians have `been slow to exploit measurement as a professional tool."¹⁰ The available evidence suggests that librarians use statistical data less often in cooperative planning - and even in administrative planning - than do personnel of other educational and service agencies Use of statistical measures by librarians in program planning appears to have occurred infrequently."¹¹

To analyze library statistics thoroughly requires a deeper understanding of modern statistics than is provided by many library school courses, especially at the master's level. In order to extract the maximum possible amount of information from a set of figures - and in order to avoid drawing superficially plausible but actually unwarranted conclusions - the professional librarian needs to know something about inferential statistics as well as descriptive statistics.

The widespread notion that only descriptive statistics (at most) needs to be taught to librarians is, unfortunately, all too clearly revealed in a recent paper¹² by J. M. Brittain. Although he argues well the case for teaching quantitative methods in library

schools, he dismisses inferential statistics with the comment that a recent text,¹³ which discusses this area, "provides too much detail about statistical techniques, many of which have limited applicability to library and information problems."

The purposes of the present paper are: (1) to suggest that library and information science are already much more dependent on the use of statistical techniques than many librarians and library educators appear to be aware; (2) to present examples of the need for a better understanding of statistics, especially inferential statistics, among librarians and information scientists; and (3) to offer evidence of what can be done in providing library school students with a more extensive introduction to statistics than is generally given them.

What are "descriptive" and "inferential" statistics? The methods of descriptive statistics serve the purpose of describing measurable characteristics of some set of entities (e.g., people, libraries, circulation figures, periodical prices). Among the characteristics frequently used are the total, counts and percentages in subsets, the median, the arithmetic and geometric means, the range, and the standard deviation: e.g., median number of books written by individual authors, (arithmetic) mean daily number of journals circulated, geometric mean annual increase in the size of a collection, range of the number of reference inquiries in a day, and standard deviation of the number of patrons using the card catalog at the same time. The techniques of descriptive statistics include determining numerical values for such characteristics, summarizing them, and displaying them in tables, graphs, charts, etc.

The methods of inferential statistics center around the process of examining a sample of data about some set of entities of interest - such a set is called a "population"- and, through use of the evidence available in the sample, making an inference about some characteristic of the population. The goals are to make correct inferences, to avoid incorrect inferences, and to have a clear idea of just how likely it is that a particular inference is correct. The usual path to this goal is to make explicit a statement, called a "statistical hypothesis," concerning the population characteristic and then to apply a statistical technique to the evidence in the sample in order to reach a decision either to accept or reject the hypothesis.

A particular virtue of inferential statistics is that it calls attention to the fact that many phenomena are by nature variable, and that observed differences may often be due to nothing more than chance. For instance, if one knows that the daily circulation of a library fluctuates easily and frequently from as low as 50 to as high as 300, one is unlikely to assume that the difference between a mean daily circulation of 100 last week and 101 this week really portends an upsurge in circulation. In less obvious circumstances, however, people have been known to interpret equally meaningless differences as important.

Inferential statistics furnishes tools by which to decide whether an observed difference is "significant," in a strict technical sense: viz., that the difference is very unlikely to be due to chance. A difference that is statistically significant may or may not be of practical importance. On the other hand, a difference may seem, to the uninitiated, large enough to be important and yet be, in fact, non-significant because it could easily have been due to mere chance.

A fundamental practice in inferential statistics is to formulate a hypothesis - often called a "null" hypothesis - that amounts to saying, "The observed phenomenon can reasonably be attributed to chance, i.e., is not significant," and then to examine the data in order to decide whether this hypothesis can be rejected. Often the investigator hopes to be able to reject the null hypothesis because he or she favors an explanation other than chance for the phenomenon.

The first purpose of this paper is to indicate that library and information science already make considerable use of statistical techniques, both in research and in practice. Of special interest is the current level of use of the techniques of inferential statistics. From an on-going study of trends in the use of statistics in library and information science, data are available on the use of statistics in articles in journals in these fields during 1975.

Before the data are presented, some definitions are necessary. Journals often contain not only articles but also editorials, news items, obituaries, etc. In this study the basic criteria for deciding what was an article were (1) identification as an article by the editor in the table of contents or (2) presence of an author's name along with the title in the table of contents. Regular columns (i.e., regularly appearing features) were not taken to be articles; neither were editorials, news stories, and letters to the editor. Doubtful items were counted as articles if they had at least one reference.

Inevitably, some decisions had to be made rather arbitrarily, but at least all such decisions were made by just one person, and the author made them before examining an item to see whether it contained statistics or not.

An article was counted as using statistics (1) if it mentioned quantitative data for the purpose of comparison (e.g., of one institution against another, one situation against another, or one time against another), or (2) if the quantitative data served to distinguish an institution or situation from others of the same general type - as in "Library A has 500,000 titles" (viz., not 5,000,000 or 5,000) - provided that the distinction was essential to the main theme of the article. Articles that incidentally mentioned quantitative data for background information were not counted. For example, the sentence, "At the beginning of 1974 there were forty-eight EDCs [European Documentation Centres] in the United Kingdom, mostly located within university or polytechnic libraries,"¹⁴ appeared in an article which, despite its containing a wealth of information, was not counted as using statistics. Finally, vague statements such as "about 50 per cent" or "over 40,000 readers" were not treated as uses of statistics.

For the purposes of this study, an article was deemed to have used inferential - as distinguished from descriptive - statistics if it employed such techniques as: analysis of variance; chi-square test of association; chi-square test of goodness-of-fit; confidence intervals for population parameters; F-test for the comparison of variances; t-test of the difference of means; or comparable techniques, either parametric (i.e., based on properties of the normal, or Gaussian, distribution) or non-parametric.¹⁵ With perhaps a slight stretch of the definition, an article that used correlation or regression was counted as an instance of inferential statistics even if the author failed to state explicitly a hypothesis about the population correlation coefficient. This does not mean, however, that articles containing loose statements about the "correlation" of two phenomena were counted. For an article to be included in the inferential statistics category, the author had to show clearly that he or she was referring to a standard correlation coefficient (e.g., Pearson, Kendall, Spearman), not merely a subjectively judged tendency toward similar behavior.

In the study all articles appearing in 1975 in 36 journals in the fields of library and information science were examined. These journals were all the national and international scholarly journals in the fields of library and information science that (1) were available in the library of the University of Texas at Austin and (2) were in English or French. Regional American journals (e.g., state library association journals) and journals in other languages were excluded.

The journals examined constitute nearly all of the population of scholarly journals in English in library and information science, plus two journals in French. One of the journals, *Scientific and Technical Information Processing*, consists of translations of "selected major articles" from *Nauchno-Tekhnicheskavalnformatsiya, Seriya I*, and thus represents some of the literature in Russian. It seems fair to consider the journals in the study as constituting a substantial portion of all the scholarly journal literature in library and information science for 1975, and a very substantial portion of the English-language subset of that literature. The year 1975 was used because it was the most recent year for which, owing to delays in the bindery, the journal literature was conveniently accessible.

The basic data from the study appear in Table 1. The total number of articles examined was 1,157, of which 522, or 45.1 per cent made at least one use of statistics that was considered worth counting according to the criteria explained earlier. Every journal that was examined turned out to have at least one article with statistics. Some 34, or 2.9 per cent of all the articles, made at least one count-worthy use of inferential statistics, an over-all average of nearly one such article per journal. These 34 constituted 6.5 per cent of the articles that used statistics.

TABLE 1. *Journals and Numbers of Articles Using Descriptive and Inferential Statistics in 1975*

Journal	Total number of articles in 1975	Number of articles using statistics	Number of articles using inferential statistics
<i>Aslib Proceedings</i>	50	12	1
<i>Australian Library Journal</i>	67	27	1
<i>Bulletin des Bibliothèques de France</i>	26	18	0
<i>Bulletin of the Medical Library Association</i>	47	30	1
<i>Canadian Library Journal</i>	48	16	0

<i>College and Research Libraries</i>	44	23	4
<i>Documentation et Bibliothèques</i>	25	10	0
<i>Drexel Library Quarterly</i>	30	6	0
<i>IFLA Journal</i>	12	2	0
<i>Information Processing and Management</i>	23	12	1
<i>International Forum on Information and Documentation</i>	5	1	0
<i>International Library Review</i>	54	36	0
<i>Journal of Academic Librarianship</i>	20	8	0
<i>Journal of the American Society for Information Science</i>	45	26	8
<i>Journal of Documentation</i>	22	16	2
<i>Journal of Education for Librarianship</i>	14	8	1
<i>Journal of Librarianship</i>	21	13	4
<i>Journal of Library Automation</i>	14	11	0
<i>Journal of Library History</i>	24	8	0
<i>Law Library Journal</i>	24	11	0
<i>Library Association Record</i>	60	21	0
<i>Library History</i>	3	1	0
<i>Library Journal</i>	53	20	0
<i>Library Quarterly</i>	23	12	3
<i>Library Resources and Technical Services</i>	32	13	1
<i>Library Trends</i>	55	36	0
<i>Libri</i>	27	14	1
<i>Microform Review</i>	10	6	0
<i>Research in Librarianship (only issues 28 and 29 were available for the study)</i>	4	4	1
<i>RQ</i>	26	7	0
<i>School Media Quarterly</i>	22	7	3

<i>Scientific and Technical Information Processing</i>	48	23	1
<i>SLJ/School Library Journal</i>	32	3	0
<i>Special Libraries</i>	67	30	1
<i>Unesco Bulletin for Libraries</i>	34	18	0
<i>Wilson Library Bulletin</i>	42	13	0
Totals	1,157	522	34

In Table 2 the journals are arranged in order of percentage of articles making at least one use of statistics. That the top-ranked journal, *Research in Librarianship*, appears to be an outlier may be due to the fact that only two issues were available for the study, containing only four articles altogether. Of the 36 journals studied, 15, or 41.7 per cent, were found to have count-worthy uses of statistics in over half of their articles. The smallest percentage of statistical articles in any journal was 9.4 per cent.

TABLE 2. Percentages, by Journal, of Articles Using Descriptive and Inferential Statistics in 1975

Journal	Percent of all articles that used statistics	Percent of all articles that used inferential statistics	Percent of statistical articles that used inferential statistics
<i>Research in Librarianship</i>	100.0	25.0	25.0
<i>Journal of Library Automation</i>	78.6	0	0
<i>Journal of Documentation</i>	72.7	9.1	12.5
<i>Bulletin des Bibliothèques de France</i>	69.2	0	0
<i>International Library Review</i>	66.7	0	0
<i>Library Trends</i>	65.4	0	0
<i>Bulletin of the Medical Library Association</i>	63.8	2.1	3.3
<i>Journal of Librarianship</i>	61.9	19.0	30.8
<i>Microform Review</i>	60.0	0	0
<i>Journal of the American Society for Information Science</i>	57.8	17.8	30.8
<i>Unesco Bulletin for Libraries</i>	52.9	0	0
<i>College and Research Libraries</i>	52.3	9.1	17.4
<i>Information Processing and Management</i>	52.2	4.3	8.3
<i>Library Quarterly</i>	52.2	13.0	25.0
<i>Libri</i>	51.8	3.7	7.1

<i>Scientific and Technical Information Processing</i>	47.9	2.1	4.3
<i>Law Library Journal</i>	45.8	0	0
<i>Special Libraries</i>	44.8	1.5	3.3
<i>Journal of Education for Librarianship</i>	44.4	5.6	12.5
<i>Library Resources and Technical Services</i>	24	8	0
<i>Australian Library Journal</i>	40.3	1.5	3.7
<i>Documentation et Bibliothèques</i>	40.0	0	0
<i>Journal of Academic Librarianship</i>	40.0	0	0
<i>Library Journal</i>	37.7	0	0
<i>Library Association Record</i>	35.0	0	0
<i>Canadian Library Journal</i>	33.3	0	0
<i>Journal of Library History</i>	33.3	0	0
<i>Library History</i>	33.3	0	0
<i>Wilson Library Bulletin</i>	31.0	0	0
<i>RQ</i>	26.9	0	0
<i>Aslib Proceedings</i>	24.0	2.0	8.3
<i>Drexel Library Quarterly</i>	20.0	0	0
<i>International Forum on Information and Documentation</i>	20.0	0	0
<i>IFLA Journal</i>	16.7	0	0
<i>SLJ/School Library Journal</i>	9.4	0	0
Medians of all values	44.6	0.0	0.0
Medians of non-zero values	44.6	5.0	10.4

Almost half of the journals contained at least one article that used inferential statistics. Table 3 presents the 16 journals that did so. The arrangement is in order of percentage of inferential statistics articles with respect to the set of articles using either descriptive or inferential statistics. (That the *School Media Quarterly* ranks first in the percentage of statistical articles that used inferential statistics perhaps stems from the fact that American graduate students who major in education often receive much training in statistics before occupying positions in such places as school media centers.)

TABLE 3. *Journals with Articles Using Inferential Statistics in 1975*

Journal	Percent of statistical articles that used inferential statistics	Number of articles using statistics	Number of articles using inferential statistics
<i>School Media Quarterly</i>	42.9	7	3
<i>Journal of the American Society for Information Science</i>	30.8	26	8
<i>Journal of Librarianship</i>	30.8	13	4
<i>Library Quarterly</i>	25.0	12	3
<i>Research in Librarianship</i>	25.0	4	1
<i>College and Research Libraries</i>	17.4	23	4
<i>Journal of Documentation</i>	12.5	16	2
<i>Journal of Education for Librarianship</i>	12.5	8	1
<i>Aslib Proceedings</i>	8.3	12	1
<i>Information Processing and Management</i>	8.3	12	1
<i>Library Resources and Technical Services</i>	7.7	13	1
<i>Libri</i>	7.1	14	1
<i>Scientific and Technical Information Processing</i>	4.3	23	1
<i>Australian Library Journal</i>	3.7	27	1
<i>Bulletin of the Medical Library Association</i>	3.3	30	1
<i>Special Libraries</i>	3.3	30	1

The primary concern of this paper is the teaching of descriptive and inferential statistics to library school students. What implications for teaching are found in the data on the use of statistical methods in the 1975 journal literature of library and information science?

First, the inferential approach leads one to the recognition that the journal literature which was published in 1975 and examined in the study is only a sample of recent research and practice. As a sample, however, it is large enough to permit one to believe that the proportions observed are fairly close to those that would be found if one examined the entire population of recent journal literature in these fields. (It is expected that the on-going study that furnished the data will find that the proportions of use of statistics in general and of inferential statistics in particular have increased in the literature of library and information science during recent decades; but at the moment, that remains an untested hypothesis.)

Second, one must recognize, especially from the data in Table 2, that even a casual reading of this literature demands an acquaintance with descriptive statistics. Even in the least statistically oriented journal, nearly one article out of ten used statistics.

Third, the data show that professionals working in library science and information science unquestionably need to understand basic inferential statistics in order to follow the literature of these fields. From Table 3 it is easy to see that the *Journal of the American Society for Information Science* had the highest number of inferential statistics articles, 8, and that the *Journal of Librarianship* and *College and Research Libraries* tied for second place with four such articles apiece. Placing third were the *Library Quarterly* and the *School Media Quarterly*, both with three inferential statistics articles; the *Journal of Documentation*, with two, was next. Journals of the caliber of these and the others in Table 3 simply cannot be ignored by the professional librarian or information scientist, and in these journals the professional is going to continue to encounter articles that he or she cannot fully understand without a knowledge of inferential statistics.

Since professionals in library and information science need an understanding of descriptive and inferential statistics, library school students need to be taught these subjects. As future professionals, they must be enabled to cope with publications that go beyond what B. C. Vickery¹⁶ calls the qualitative description mode of investigation into what he labels as quantitative and evaluative description and, even further, into correlative studies and studies that involve causal analysis. Whether or not students and professionals perform such studies themselves, they must be able to understand, and appreciate the importance of, such studies.

There is still another way of interpreting the data. Although many of the articles counted as using statistics did so in minor ways, a considerable number of articles contained large quantities of descriptive statistics. Many of these were detailed presentations of results from surveys or records of user behavior (e.g., studies of interlibrary loan requests with data on age of journal or book sought, category of requestor - physician, dentist, student, social worker, medical technician, - language of item requested, source of reference.) Many offered a richness of data that cried out for inferential analysis (e.g., do physicians who are primarily in practice really tend to request articles that are more recent than those requested by physicians who are primarily educators?).

Brittain asserts that the "needs of practising librarians for skills [in research and statistics] are relatively modest . . ."¹⁷ To the degree that this is true, it is so because practicing librarians and information scientists set their sights too low and fail to realize how much better they could practice their profession if they possessed statistical skills and were alert for opportunities to employ them. Burns, having observed that "libraries have never come to grips adequately with the problems which arise in gathering statistical information about themselves or their operations,"¹⁸ goes on to present a long list of questions to which library and information system managers need answers but almost always lack them. Why are the answers missing? Because the generally inadequate statistical knowledge of professional librarians and information scientists causes them to miss opportunities to gather and analyze data that could yield the answers.

The percentage of use of descriptive statistics in journal articles in library and information science can be viewed as a measure of the readiness of authors in these fields to employ quantitative arguments, to appeal "to the numbers" for support for their theses. The percentage of use of inferential statistics can be viewed as a measure of the sophistication of authors in these fields in employing quantitative arguments. The gap between the percentage (45.1) of articles in 1975 that used statistics of any kind and the percentage (2.9) that used inferential statistics indicates the great potential for improvement in the use of quantitative arguments.

The study uncovered example after example of descriptive summaries of data on which inferential analyses might have been performed. Such analyses might have led - surely would have led in some cases - to a deeper understanding of the phenomena being studied.

One clarification that can result from inferential analysis is a recognition that apparently minor differences are actually significant (in the statistical sense) and may be of practical importance. Failure to use inferential analysis can thus result in missing an opportunity for clarification.

Probably more common is the mistake of assuming that an observed difference is significant when inferential analysis would have shown it to be easily explained by chance. An example which the author has used in the classroom for several years comes from an article¹⁹ comparing men and women head librarians in American colleges. One of the tables in that article contains the information presented here as Table 4.

TABLE 4. *Portion of College Budget Allotted to Library vs. Sex of Head of Library*

Portion of college budget allotted to the library	Number of librarians whose libraries were allotted the stated percentage of the college budget	
	Male head librarians	Female head librarians
Under 2%	0	6
2% - 4%	78	60
4% - 6%	92	109
Over 6%	19	18

Referring to the data in Table 4, the article says, "It would appear that the ladies are slightly better at getting the money than are the men." Although this statement can be argued to be correct if interpreted literally, it certainly might leave a hurried reader with the erroneous impression that the article had proved women head librarians in general to be more successful than men head librarians at "getting the money." In fact, by applying the chi-square test of association to the data in Table 4, one obtains 4.35 as the observed chi-square score. With 3 degrees of freedom, this score is only slightly over half of the critical value of 7.82 needed to reject, at the 5 per cent level of significance, the null hypothesis of no association between a library's share of the college budget and the head librarian's sex. One must therefore recognize that the numbers of male and female head librarians that turned up in the author's sample might just as easily have seemed to show a slight tendency for men to be better than women at "getting the money." The observed data are simply inconclusive.

A clear-cut example of an author's reaching a conclusion that was not supported by his own evidence is found in a paper²⁰ surveying users of the technical library of a company. One of the statements in the paper is: "Significant differences between directors [of departments of the company] and non-directors were revealed in many responses On their reason for use of the library [presented here as Table 5], 44% of the directors use it for keeping up on current information compared with 21% for non-directors. This director/non-director split has serious implications for the library and is *the* most important finding."

TABLE 5. *Reasons for Using a Company Library vs. Status of User*

Reason for Use of Library	Consensus [Totals]	Director	Non-Director
Work in quiet place	0	0	0
Literature search	1 (3%)	0	1 (7%)
Keep up on current literature	11 (34%)	8 (44%)	3 (21%)
Get answer to question	13 (41%)	8 (44%)	5 (36%)
Obtain citation	7 (22%)	2 (12%)	5 (36%)

What happens when one examines these data by the methods of inferential statistics? Using the Kolmogorov-Smirnov two-sample test of goodness-of-fit,²¹ one finds that the observed value, 0.24, of the test statistic is much less than the critical value, 0.48, which (with samples of 14 and 18, respectively) must be exceeded for rejection of the null hypothesis at the 5 per cent level of significance. The general null hypothesis of the Kolmogorov-Smirnov test is, H_0 : The two populations from which the samples were drawn have the same distribution function. In this example the null hypothesis may be particularized thus: The various reasons for using the library are shared, in the same proportions, by directors and non-directors.

In short, when one examines the data of the example from the viewpoint of inferential statistics, one realizes that the apparent difference between directors and non-directors in their reasons for using the library is probably just the kind of difference that arises

by chance alone out of typical vagaries of the process of sampling. The evidence simply does not justify the reported conclusion that directors and non-directors generally differ in their reasons for using the library.

One need not be versed in the Kolmogorov-Smirnov test to check the conclusion the article draws from Table 5. Though the Kolmogorov-Smirnov procedure is preferable for dealing with Table 5, the better known chi-square test of association could be used. An appropriate way of using it entails ignoring the "work in quiet place" category and merging the "literature search" and "obtain citation" categories to avoid expected frequencies less than one.²² The result is a chi-square score of 4.54 with 2 degrees of freedom, well under the critical value of 5.99 needed for rejection, at the 5 per cent level, of the null hypothesis of no association between being a director or non-director and reasons for using the library.

Both the chi-square and the Kolmogorov-Smirnov procedures show that the article's "most important finding" is not supported by its data. The point is that its author failed to recognize the possibility that his data were inconclusive and to test for that possibility. Now, all students in an elementary course in inferential statistics will work at least a few problems using the chi-square test of association. The author would argue that any such student experiences enough of the vagaries of small samples to have asked himself or herself whether the frequencies in Table 5 really supported the conclusion drawn in this example, and that any such student would thus have avoided the error.

Even an elementary understanding of the concepts of inferential statistics would have saved the author of this example from stating in print- as "*the* most important finding" of his study - a conclusion which his own data belie. But more important, such an understanding would have saved his readers from acquiring as fact an unsubstantiated, and quite possibly false, conclusion.

A final example. While the author was writing this paper, a former student dropped in to visit. During the conversation he casually mentioned that he had just found a potentially embarrassing misuse of statistics in a request for a grant currently being drafted in the library where he now works. Because he knew enough about statistics, he not only kept his institution from being embarrassed by the misuse but also quite possibly prevented the proposal itself from being considered unfavorably by the prospective funding agency because of the error. In short, his knowledge of statistics may well mean thousands of dollars to his library in this instance alone.

The problems of making mistakes and missing opportunities to probe more deeply into phenomena constitute the main argument for teaching inferential statistics in library schools. One should not overlook the additional argument of enabling students to understand the professional literature better.

But can library school students learn inferential statistics? The present paper is arguing that the answer is "Yes," and this answer is supported by experience in teaching inferential statistics in the Graduate School of Library Science, University of Texas at Austin.

For the past six years the course, Research in Library Science, has included an introduction to descriptive and inferential statistics. Other parts of the three-semester-hour course are an introduction to scientific research, training in the preparation of research proposals, training in survey methods, and brief treatments of other areas of library-related research, viz., content analysis, stylostistics, historical research (with emphasis on cliometrics, the newly developing area of quantitative methods in historical research), and analytic bibliography.

This course appears quite similar in content to the quantitative methods course at Loughborough University that Brittain²³ outlines, except that it places less emphasis on survey methods and much more emphasis on statistics. These differences in emphasis stem not only from the case for teaching inferential statistics presented in this paper, but also from the belief that students (and professionals) can learn survey methods on their own much more easily than they can learn statistics on their own.

To guide the autodidact in survey methods, there is an abundance of good textbooks. The appendix to this paper offers some suggestions.

The would-be autodidact in statistics, however, faces greater difficulties. There is simply no satisfactory textbook in English²⁴ on statistics designed specifically for librarians and information scientists, although H. Goldhor's book²⁵ on research provides a useful introduction. A recent effort toward such a text²⁶ turned out, unfortunately, to be "replete with errors"²⁷ that make it unsuitable for the statistical novice. For textbooks in statistics, the Texas course has relied on Goldhor's book and on introductory texts for students in psychology and sociology (for details, see the appendix).

The intrinsic difficulties of statistics and the lack of suitable texts aimed at librarians and information scientists would suffice as reasons for emphasizing statistics over survey methods in library school courses in research or quantitative methods. But there is still another reason: It is now possible to teach students computer-assisted statistics. Computer programs can handle the tedious arithmetic that used to burden students and statisticians. The student, freed from arithmetic toil, can concentrate on the concepts of statistics instead of the calculations. Furthermore, computer programs can easily handle problems large enough and realistic enough to provide the student additional incentives for learning. Practice with such programs helps students to become aware of how easy it really is nowadays to get answers to statistical problems. Clearly, it will ordinarily be much easier (for administrative, rather than technical, reasons) to provide instruction in computer-assisted statistics to organized classes in universities than to autodidacts.

The Texas course in research in library science integrates the learning of statistical concepts with practice in using several statistical program packages,²⁸ including BMD,²⁹ OMNITAB II,³⁰ and SPSS.³¹ Barely have the students learned what the arithmetic mean is before they learn how to make OMNITAB calculate means for them. By the end of the course the students are using the computer to perform regression analyses and to draw data points and trend lines on a computer plotter. Along the way the students learn the concepts of inferential statistics and some of its principal techniques (analysis of variance, chi-square, confidence intervals, correlation, regression, and t-test) by working library-oriented problems with computer assistance.

The total cost of computer time and supplies for the problems is about \$15 per student. For this modest expenditure by the university the students acquire not only a knowledge of inferential statistics but also a good deal of practice in using computers interactively and in batch-mode. The practice is effective in overcoming the qualms about using computers with which many students begin the course. Furthermore, the statistical program packages used in the course are widely available. The students' attention is called to the fact that many of them will be working for academic institutions, municipal governments, school systems, and companies with computer systems on which these or comparable statistical program packages are or can be installed.

Lest anyone think that students emerge from the course able to work statistical problems only with the aid of a computer, it should be added that manual methods are not neglected. However, arithmetic for the sake of arithmetic is avoided, and students are encouraged to use personal electronic calculators. Quite inexpensive models (under \$30) now offer direct calculations of mean and standard deviation. For student use Texas provides three electronic calculators - Hewlett-Packard models 67 and 45 and a Commodore model S-61 ("Statistician") - and the students are required to learn to use all three.

Experience indicates that a course in research or quantitative methods ought to come early, rather than late, in the education of the fledgling librarian or information scientist. At Texas the research methods course used to be scheduled for the student's terminal semester, as one of the final steps in preparing the student to enter the profession. But at the suggestion of students the course has been moved to early in the curriculum. Thus the students are able to use what they learn in the research course in their later courses, especially in connection with their reading of the professional literature for other courses. This recognition by the students themselves of the importance of statistical methods is further evidence of the indispensability of these methods for the person who wants to function as a professional in library and information science.

Over 700 students have completed this course in research in library science during the past six years. Their success in the course proves that library school students can learn both descriptive and inferential statistics.

Summary. A study of the journal literature of library and information science showed that in 1975 about 45 per cent of all the articles made some use of statistics and about 3 per cent of all the articles used inferential statistics. The former percentage was interpreted as, *prima facie*, a call for the educating of library school students in descriptive statistics. Although the latter percentage is low, articles using inferential statistics appear with higher relative frequencies in the leading journals used by serious professionals in library and information science. The low percentage of inferential statistics articles was also placed in the context of mistakes in the analysis of quantitative data and missed opportunities for more penetrating analyses of such data. It was argued that such mistakes and missed opportunities, along with the use of inferential statistics in the leading professional journals, call for the educating of library school students in inferential statistics. A description was given of a course in research methods that includes descriptive and inferential statistics and provides students with experience in using computer program packages to handle statistical data. The successful completion of this course by several hundred library school students provides the evidence that such students can learn descriptive and inferential statistics.

APPENDIX

Recommended Works on Survey Methods

Berdie, D. R., and Anderson, J. F.: *Questionnaires: Design and Use*. Metuchen, New Jersey, Scarecrow, 1974.
Hyman, H. H., et al.: *Interviewing in Social Research*. Chicago, University of Chicago Press, 1954.
Line, M. B.: *Library Surveys*. London, Clive Bingley, 1967.
Moser, C. A.: *Survey Methods in Social Investigation*. London, Heinemann, 1958.
Oppenheim, A. N.: *Questionnaire Design and Attitude Measurement*. New York, Basic Books, 1966.
Tauber, M. F., and Stephens, I. R., eds.: *Library Surveys*. New York, Columbia University Press, 1967.
Wheeler, M.: *Lies, Damn Lies, and Statistics: The Manipulation of Public Opinion in America*. New York, Liveright, 1976.

Comments on Textbooks in Statistics

Beginning with the spring semester of academic year 1977-78, the textbooks in statistics for the course, Research in Library Science, at the University of Texas at Austin, are:

Wallis, W. A., and Roberts, H. V.: *Statistics: A New Approach*. Glencoe, Illinois, Free Press, 1956.
Bruning, J. L., and Kintz, B. L.: *Computational Handbook of Statistics*. 2nd ed. Glenview, Illinois, Scott, Foresman, 1977.

The previously used textbook for the course was:

Hardyck, C. D., and Petrinovich, L. F.: *Introduction to Statistics for the Behavioral Sciences*. 2nd ed. Philadelphia, W. B. Saunders, 1976. The approach in this work is sound; however, the first printing of the second edition suffers from an excessive number of typographical errors (this author can provide a list of errata).

A text in use earlier for the course has just been published in a new edition:

Young, R. K., and Veldman, D. J.: *Introductory Statistics for the Behavioral Sciences*. 3rd ed. New York, Holt, Rinehart and Winston, 1977. Changes in the new edition appear to have increased the book's suitability for library school students.

References and Notes

1. Palmer, D. C., ed.: *Planning for a Nationwide System of Library Statistics*. Washington, D.C., U.S. National Center for Educational Statistics, 1970.
2. Childers, T. A.: Statistics That Describe Libraries and Library Service. In: Voigt, M. J., ed.: *Advances in Librarianship*. New York, Academic Press, 1975. Vol. 5, pp. 107-122.
3. Anders, M. E.: Statistical Information as a Basis for Cooperative Planning. *Library Trends*, 24:229-244, Oct. 1975.
4. Burns, R. W., Jr.: An Empirical Rationale for the Accumulation of Statistical Information. *Library Resources and Technical Services*, 18:253-258, Summer, 1974.
5. Drake, M. A.: Forecasting Academic Library Growth. *College and Research Libraries*, 37:53-59, Jan. 1976.
6. Leech, S.: You Can't Add Apples and Oranges. *Kentucky Library Association Bulletin*, 39, no. 4: 24-27, special issue, 1975.
7. Williams, W. H.: How to Prove Your Library's Worth to Its Community. *Wyoming Library Roundup*, 31, no. 2:7-8, June, 1976.
8. See, for example: Slanker, B. O.: Developing LIBGIS with State Participation. In: Miele, M., ed.: *Bowker Annual of Library and Book Trade Information*. 19th ed. New York, R. R. Bowker, 1974, pp. 225-228.
9. Leech, ref. 6.
10. Purdy, G. F.: Professional Overview. In: Palmer, ref. 1, pp. 25-29.
11. Anders, ref. 3.

12. Brittain, J. M.: Teaching Quantitative Methods in Library Schools. *Journal of Librarianship*, 9:108-119, April, 1977.
13. Simpson, I. S.: *Basic Statistics for Librarians*. London, Clive Bingley, 1975.
14. Hopkins, M.: The Bibliographical Resources of European Documentation Centres. *Journal of Librarianship*, 7:84-99, April, 1975.
15. Concise descriptions of many parametric and non-parametric techniques of inferential statistics may be found in: Roscoe, J. T.: *Fundamental Research Statistics for the Behavioral Sciences*. 2nd ed. New York, Holt, Rinehart and Winston, 1975.
16. Vickery, B. C.: Academic Research in Library and Information Studies. *Journal of Librarianship*, 7:153-160, July, 1975.
17. Brittain, ref. 12.
18. Burns, ref. 4.
19. Blankenship, W. C.: Head Librarians: How Many Men? How Many Women? *College and Research Libraries*, 28:41-48, Jan. 1967.
20. Fatcheric, J. P.: Survey of Users of a Medium-Sized Technical Library. *Special Libraries*, 66:245-251, May/June, 1975.
21. See, for example: Conover, W. J.: *Practical Nonparametric Statistics*. New York, Wiley, 1971.
22. For a discussion of what minimum value of expected frequencies should be used in the chi-square procedure, see pp. 235 and 241 of: Snedecor, G. W., and Cochran, W. G.: *Statistical Methods*. 6th ed. Ames, Iowa, Iowa State University Press, 1967.
23. Brittain, ref. 12.
24. A recent work in German deserves special mention: Stock, K. F.: *Grundlagen und Praxis der Bibliotheksstatistik*, Pullach bei München, Verlag Dokumentation, 1974. A reference handbook rather than a textbook, this work contains an excellent overview of descriptive and inferential statistics, discusses numerous possible applications, and provides many examples of analyses of library statistics.
25. Goldhor, H.: *An Introduction to Scientific Research in Librarianship*. Urbana, Graduate School of Library Science, University of Illinois, 1972.
26. Simpson, ref. 13.
27. Norton, H. W.: Review of Basic Statistics for Librarians, by I. S. Simpson. *Newsletter on Library Research*, no. 18.8-10, Dec. 1976.
28. Wyllys, R. E.: Instructional Use of Statistical Program Packages: BMD, IMP, OMNITAB II, and SPSS. In: Hogben, D., and Fife, D. W., eds.: *Computer Science and Statistics: Tenth Annual Symposium on the Interface*, pp. 265-270. (National Bureau of Standards Special Publication 503.) Washington, D.C., National Bureau of Standards, 1978.
29. Dixon, W. J., ed.: *BMDP: Biomedical Computer Programs*. Berkeley, California, University of California Press, 1975.
30. Hogben, D., Peavy, S. T., and Varner, R.N.: *OMNITAB II User's Reference Manual*. (NBS Technical Note 552.) Washington, D.C., National Bureau of Standards, 1971.
31. Nie, N. H., et al.: *SPSS: Statistical Package for the Social Sciences*. 2nd ed. New York, McGraw-Hill, 1975.