Statistical Assumption-Making in Library Collection Assessment: Peccadilloes and Pitfalls

Richard Hacken
hacken@byu.edu

Follow this and additional works at: https://scholarsarchive.byu.edu/facpub

Part of the Collection Development and Management Commons

Original Publication Citation

BYU ScholarsArchive Citation
https://scholarsarchive.byu.edu/facpub/1207

This Peer-Reviewed Article is brought to you for free and open access by BYU ScholarsArchive. It has been accepted for inclusion in All Faculty Publications by an authorized administrator of BYU ScholarsArchive. For more information, please contact scholarsarchive@byu.edu, ellen_amatangelo@byu.edu.
ABSTRACT. Assessing library collections in the Semiconductor Age necessarily involves a heavy use of quantitative data. The assumptions made during the process of gathering, manipulating, and reporting library statistics may or may not be valid ones. Objective and vigilant scrutiny, therefore, can make the difference between an assessment that adds to a greater knowledge of the collection and one that only adds greater bulk to The File. Among the areas affected by statistical assumptions are (in lay terms): the sample, the survey, the percentage, the average, and the degree of accuracy.

The librarian of the eighties cannot ignore "statistics," especially after having grown up in a world dominated by "9 out of 10 doctors," "47% fewer cavities," and "3.55 grade point average." The authors of Statistical Methods for Librarians note this.

As a consumer of research, the librarians finds the literature of his or her field increasingly statistical. The journals, books, and other resources from relevant fields, such as business administration, sociology, or political science, are likely to be even more quantitative. As active researchers, either as resource persons or as principal investigators, librarians are virtually bound to handle numerical data.¹

Number-crunching has become so respectable in our society and so facilitated by our silicon-based colleagues that it seems natural and right to transpose our books and journals, patrons and bibliographic records into manageable digits—and then to form decisions based on those numeric configurations. The purpose of this article is neither

¹ Richard D. Hacken, PhD, is European Studies Bibliographer, Harold B. Lee Library, Brigham Young University, Provo, Utah 84602.

Statistical Assumption-Making in Library Collection Assessment: Peccadilloes and Pitfalls

Richard D. Hacken
to assert cynically that such quantification cannot be done nor to question the legitimate results of statistically-based collection assessment; rather, it is to hoist a warning flag against the sloppy and unconsidered assumptions that lead to faulty conclusions.

In a university library, false collection assessment decisions can come back to haunt later generations of users. A written and formalized assessment—whether accurate or flawed—may not be questioned or reassessed for years, during which time the collection is growing in the direction prescribed in the assessment. On a negative note, it may not be pure fantasy to suppose that someday a library might find itself involved in a lawsuit for "failure to provide relevant educational support," just as school districts and universities have been hauled before the bar for "failure to educate." I have recently received in the mail an insurance company's offer to provide liability coverage up to half a million dollars should I have to battle educational neglect allegations up to and including "improper methods employed in instruction, counseling, research design, etc." or "negative consequences in the implementation of the recommendations of research studies." This would suggest culpability not only for one's own faulty research, but also for being lured down the garden path by implementing the faulty research of others. While I do not take the legal dangers seriously enough to enroll in this liability plan, I do take them seriously enough to remind myself how soberly and objectively research in collection assessment should be carried out.

A recent letter to the editor of College and Research Libraries illustrates the concern for relevant and reliable interpretations of library data. Although the letter writing impugns an assessment that deals with librarians rather than collections, the flaws in research design that come into question could invalidate any type of assessment survey:

The sample is never shown to be a valid representation of academic librarians or a sub-group of academic librarians and thus, is not generalizable (especially with a 52% response rate), a copy of the questionnaire is not available as an appendix for the reader, key definitions (such as what exactly constitutes an "article") are not provided, huge assumptions are made as to participants' interpretation of questions, no explanations of the limitations and weaknesses of the study or its findings are offered to the reader, statistical techniques are poorly utilized, and there are no indications of the reliability or validity of the data reported.

Assuming that the letter writer's complaints are justified, there is more at stake than just self-delusion, for librarians—as surely as other members of our society—are encouraged to remember, repeat and build upon the statistics offered by others:

Unfortunately, this type of research simply reinforces the notion to other academic librarians that such articles are "good research" and worse, someone else, with even less knowledge conducting and interpreting research, will use if for an inappropriate reason and to justify or support a position, decision, or activity, for which it is either invalid or inappropriate.

We might further assume, charitably, that disfiguring errors found in library data interpretations most frequently occur from a lack of training, experience, or concentration rather than from a lack of intelligence or scruples. It would seem logical, then, that training ourselves in the appropriate statistical methods, gaining experience, and vigilantly concentrating on the task (while applying generous helpings of common sense) would do much to help us assess collections relevantly and accurately.

"There is something fascinating about science," Mark Twain once wrote, "One gets such wholesale returns of conjecture out of such a trifling investment of fact." Such a case of wholesale conjecture arising from a trifling investment of bibliographic fact is the tongue-in-cheek parody of two scholars who conclude that if the publication, distribution and accumulation of the National Geographical Magazine continues at the present rate (as determined through studies of the dimensions, weight and density of ten randomly selected issues), then the North American continental land mass will eventually sink under the load, flooding major coastal cities. The main argument between the two "scholars" deals with how soon this will occur. Such might be said to overstep the bounds of sober collection assessment. Still, we may find ourselves tempted to reach frighteningly similar—but less grandiose—collection conclusions using homemade "data enrichment methods." Are we really seeking to choose details that may be "skulking almost unnoticed in the raw data," or is our goal that proposed in satirical tones by a member of the Operations Evaluation Group at MIT?:
The ultimate objective, complete freedom from the inconvenience and embarrassment of experimental results, still lies unattained before us.8

One major assumption made by those who count and compare library holdings, patron responses, or bibliographic records is that a connection can be found between the quality they seek to improve and the quantities they are actually measuring. It is absolutely imperative to realize that an adversary relationship does not need to exist between numeric values and collection values.

Most of us shy away from anything called "quantitative" because we believe that quantity is the enemy of quality. If the quantity of cars increases, we are sure the quality will decrease. If the number of students goes up, the quality is supposed to go down, and so on. I will not tarry today to challenge this axiom, which you learned at your mother’s knee. I will only ask you to consider the possibility that there is no such antipathy between quantitative analysis and qualitative analysis. On the contrary, each gives meaning to the other.9

 Granted, there is a difference between quality and quantity. But in taking the full measure of a library or part of a library, both need to be taken under advisement in a complementary study. Just as a qualitative measure telling us that seawater contains gold is not fully useful until we have a quantitative measure of gold in parts per million, likewise a qualitative measure telling us that Shakespeare is represented on the shelves of both the Folger Library and the Lincoln Park Elementary School’s “media center” is incomplete until we know some relevant numbers. Evaluating the significance of varying degrees of quantity, and tying those numbers to a measure of how well a library is fulfilling its function becomes the task of a collection assessor: “Quantitative analysis is never a substitute for qualitative; it is, rather, a further step in the same process.”10

There is hardly a field of assessment study in which we cannot gain new insights merely by thinking of the problem in quantitative terms. Consider, for instance, the question of the relative place of English-language and foreign vernacular materials within a given subject area. All of us recognize that we have a legitimate interest in providing readable and relevant English materials for a North American library, that we have an equally legitimate interest in making foreign-language matter available for those who require it, and that these interest may at times conflict in the budgetary squeeze. Those who consider the question a qualitative one may well ask “Which is the higher value?”, and see no way of finding an objective answer (even if subjective ones may abound).

In this case, though, it seems to me that quantitative analysis offers a useful way of thinking about the dilemma. Instead of asking myself whether English-language materials are more or less important than foreign ones, I ask myself how many patrons will suffer if exclusively English-language materials are purchased in a subject area being considered, and how many patrons will suffer from exclusively foreign-language books.

Asking the question in this way not only opens up the possibility of quantitative analysis, but reveals other alternatives. I realize that strictly-English and strictly-foreign collections are not a research library fact of life, and that the question itself may be forced. I am able to insert other hypothetical figures into my analysis, and see how I feel about the results. If I visualize the patrons of a Russian literature reading novels in a strange Cyrillic code which they have thoroughly studied, I am able to tolerate a high number of works that are non-English. If I believe the browsers in the cookbook section, on the other hand, would be confronted with the same percentage of foreign works set in strange diacritical markings, transliterated alphabets and milliliter and gram measures, I will be prepared to accept very few non-English-language offerings. In other fields of study, an optimal mix of English and foreign books might be conceived by quantifying patrons’ needs and comparing them to present holdings. This can suggest a mix that will satisfy the greatest number of needs for the least amount of expenditure.

Thinking about the problem in this way, I find that certain problems become manageable on certain assumptions of fact. What those facts are is perhaps not as important as how I test the assumptions. I am also helped to pinpoint the differences in presupposition which divide some of my colleagues.11

The simpler types of assumptions we might introduce into various assessment procedures—either unconsciously or consciously—can be either potentially devastating (if they fly in the face of logic) or necessary and valid (if we check at every step that the assumptions still hold true). Other types of assumptions may be structurally built into the assessment procedure itself, and should be recognized as such. Yet the type of assumption made depends heavily on the type of statistics being used.

The use of “descriptive” statistics, for instance, assumes that we
have collected every bit of data possible in a certain area, and can now manipulate and illuminate the numbers.

If you have a list of the daily circulation figures for a year, you will need to perform some kind of operation to make the numbers manageable and meaningful. The process of doing so is a kind of summarizing, such as finding an average daily circulation for a year. At this descriptive level we restrict our interest and generalizations to the case or the body of data at hand.

Notice how the judicious use of descriptive statistical comparisons can illuminate a point (the reference here is to the annual book production of the German Democratic Republic):

The figure of 140 million books printed annually is much more impressive when one considers that the country has only 17 million inhabitants; the per capita production of books places the GDR in third place among the nations of the world behind the Soviet Union and Japan. Statistically seven to eight books are produced per person per year.

The last sentence in the example above, "Statistically seven to eight books are produced per person per year, " would be ambiguous and confusing if allowed to stand alone (seven to eight books produced by or for each person?). In an explanatory context, it becomes clear. The same principle holds for all levels of statistical reports: full figures are meaningless without full explanations.

The use of "inferential" statistics, on the other hand, involves taking a sampling of data and inferring general statements from the specifics.

When we have only a part of the entire body of data about one or more variables, we may employ inferential or inductive statistical procedures. In short, we may have only a sample of all daily circulation figures and want to make some general statements about the circulation.

This, obviously, can involve the heavy use of inductive assumptions.

It is clear that assumptions can be either positive or negative factors in statistical thinking; they are by nature neutral but necessary.

Assumptions only become a hindrance when they are made illogically or unnecessarily. It is equally clear that we require a great number of valid assumptions to even make it through the day as functioning human beings, let alone librarians. For instance, we assume that we will awaken to find our little world essentially the way we left it the night before; that our food, shelter, and means of transportation will be in the same shape as we left them. We assume that the road to the library will still be open and passable; that the library will still be at the same address, and that we will still be employed. Without worrying a great deal about such matters, these are some of the automatic assumptions we make that are valid in most cases; that is, they prove true.

There are a number of assumptions unique to the gathering, interpreting, and spewing of statistics. We assume, as previously expressed, that quantitative measures can say something important about quality. We assume that the methods are accurately carried out and reported. The measurements that we make are presumed to be valid; the collection, service, or other item we wish to assess, and we presume that there exists an ideal against which the findings can be weighed. We further assume that the measurement is relevant to actual needs, can answer specific questions, and displays clearly defined units.

All numbers going into a statistical calculation should be relevant and appropriate, and even if those numbers are calculated with the best available evidence, they are meaningless without an associated measuring stick. That is, we need to know the rules by which the assessment was done, so that the results can be replicated, if necessary.

We must be certain, for accuracy's sake, that no statistical results have been ignored; otherwise the results will be one-sided and skewed. The omission of relevant information is easy to do, but treacherous. In assessing the needs and problems of library users, for example, any calculations that leave out the sleepers, employees, one-day visitors, social animals, custodial help, thieves, and book mutilators will furnish only incomplete and possibly misleading information.

One area of inferential statistics in which constant and thorough testing of assumptions is required is the realm of "sampling." When the doctor takes a blood sample to check for high or low levels of this or that, he or she assumes that the miniscule amount taken is representative of your entire blood supply. If the doctor
were unable to make that assumption, it might prove necessary to drain your entire circulatory system just to make sure the composition of blood remains constant from head to toe. This would prove a hardship. In like manner, when we take a sample of books from the shelves, bibliographic records from the catalog, or patron names to survey from the list of possible library users, we have to assume that the population will be represented well by the samples we take. Unlike the doctor, however, we cannot be content to take a small cluster sample and consider it representative. We are faced with the task of making certain that the manner of sample-taking gives the highest possible probability of objectivity: "One of the more obvious sources of error is in designing samples."  

When we ask users to voice their opinions and concerns about the part of the library they use, we either poll all patrons, or assume that our sampling of users is adequate. Sampling procedures—when necessary—are assumed to be large, random and representative enough to permit validation of the study. 

As an example of "representativeness" in gathering survey information, consider the following. My university has a browsing collection of current popular materials, which—circulation records show—are borrowed 50% by students and 50% by faculty. It might, therefore, seem a warranted procedure to distribute survey questionnaires to those persons seated in the room—reading books from the collection—and to presume this to be a representative sample of borrowers. For more than one reason, however, the assumption can be considered false. First, a physical observation of the patrons in the room would suggest that almost no faculty spend any time in the browsing collection proper, but apparently remove and check out the books for use at home or in offices. Next, if it is borrowers we are interested in, then giving a survey to those seen sitting in the room is no guarantee of reaching the target group. Finally, we need to consider the impact of vastly differing loan periods to students and faculty—two weeks versus over a year—on the book availability and user satisfaction for each group. 

Obviously, the survey structure is a source of potential mischief. Careful thought has to be given to the survey design: 

A cleverly designed survey will ask crucial questions in a variety of ways to increase the probability that favorable results are obtained.  

Even with a well-structured survey, though, we have to be careful about assuming too much. A relative ranking offered by survey respondents is only a ranking, and does not provide definitive values. 

For instance, in rating reference services, the book collection, and the record-tape-film collection the user may rank the book collection first (most important), record-tape-film second, and reference services third. We know only the relative position of these services; there is no known "distance" between the first and second and the second and third choices.  

If the main concern in taking samples is that they be random enough, it is equally clear that some methodologies can be too random: 

Applying the randomization principle to experimental medicine has led to the triple-blind test: the subject does not know what he is getting, the nurse doesn't know what she is giving, and the investigator doesn't know what he is doing. Half way through the experiment, randomization is increased by a process known as turnabout—the patient administers the drug to the investigator, and the results are evaluated by a student-nurse.  

Obviously, in applying randomization to sample-gathering for the collection of library data, we cannot allow our methods and concentration to be "random" as well. 

One of the requisite traits for anyone assessing a collection is healthy skepticism—especially toward one's own unspoken assumptions. Since the potential for error exists from the earliest stages of data preparation, there is no stage at which assumptions can be left unexamined. Consider, if you will, the following (fictitious) report: 

The reference services at XY University are considered by the library administration to be a great success, since only 148 out of the 10,000 students on campus this year have complained about it. "When less than 1.5 percent of the students are discontent and more than 98.5 percent are delighted," one administrator pointed out, "as far as I'm concerned, that's a better mandate for the status quo than any politician has ever received!"
Clearly, the "sample" of reference services satisfaction at XY University is a biased one, not so much because we cannot assume the 148 complainers were in fact "discontent," but because we can by no means assume that the remaining students were "delighted" with the reference service. One good test for a collection assessor using inferential statistics (which the above example only mimics) is to ask whether there might be any alternative reasons for a given result. If we are trying to decide something, using statistical methods, we cannot overlook the possibility of chance as an explanation for the observed results.

Even when we have an inference based on evidence for which there is no exculpatory explanation, we cannot say with certainty what caused the particular numbers or evidence to appear... There may be a high correlation between the number of clergymen in a town and the number of newborn children. Without more information, we are unwilling to say the former caused the latter. The more exculpatory explanations we can eliminate, however, the more likely we are to accept this explanation of causation.

In the reference service example, there could be any number of good reasons for dissatisfied patrons' silence. We need to know how many students used the library, how many of that smaller number used the reference service, how many of those students were disappointed in the service, and what percentage of those became so disgusted that they actually complained. Finally, how many of the complaints would have been passed along to the front office? If I were in the library administration at the above (fictitious) institution, I believe I would be very anxious to know why any students at all would be dissatisfied with the reference service. That there is any quantity of alienated library users might suggest something to me about the quality of the reference function.

A statistical curiosity of special interest to the collection assessor is the percentage. We assume that percentages will give valid information about the strength or weakness of the collection (particularly when compared either longitudinally with the same collection or latitudinally with another collection of known magnitude). One area of potential misunderstanding is that of reporting collection percentages. It should go without saying that percentages are extremely misleading if the base values are too small. A classic example was the published report that 33% of the women students at Johns Hopkins University were married to professors. This figure took on less shock value when it was revealed that registration had just been opened up to females, and that only 3 women had registered the first year—of which one had married a faculty member. Similarly, it might be altogether impressive to report that a library collection in Extra-Terrestrial Cybernetics contains 80% of those works listed in the definitive bibliography on the subject. However, if only 5 works have been published, and the collection lacks the most respected and oft-quoted one, then this 80% is more suspect than if there were, 1,000 monographs on the subject, of which the subject selector has deleted 200 minor textbooks and outdated treatises. The key here is that no assumptions should be left to the reader of an assessment report; if percentages are given, then the figures that led to the calculations should also be reported. Let us look at one statement that goes beyond the problem of percentages to include other assumptions:

Seventy-five percent or more of the books that have been added to this collection since the conversion to the Library of Congress system eight years ago have clearly never been used. I have to wonder why most of them were ever acquired in the first place.

What can you as a collection assessor do with this report? The percentage figure seems to be an estimate, but of what? How clear is it which books have never been used? How many books are non-circulating and would thus have no record of transactions? Which have been taken from the stacks and returned directly to their proper place by fastidious patrons? More importantly, since the examiner is apparently referring to the most recent portion of the collection that displays Library of Congress call numbers, and has generalized about all books received in the past eight years, do we know how many of those books have only been received in the past year, and have not yet been used for that reason? How many have been added in the past two years? And yet the implication is that all these books have lain idle on the shelves for eight years. Obviously, more information and outside verification is needed to reach the conclusion that "I have to wonder why most of them were ever acquired in the first place."

In fact, clear and unequivocal definitions are an integral part of
an assessment. When we refer to the "economics collection," does this include business? commerce? banking? history of economics? Or is it defined by a certain range of call numbers? For the assessment to have any value, the exact boundaries must be set in concrete and clearly outlined in the assessment report. As long as a "collection" is clearly defined, its holdings can be compared in later years, and apples are not being compared to pomegranates. Statements such as the following must supply definitions to be considered objective: "After pumping money into the Computer Science collection for 15 years, it is now more anemic than it was then." What is the definition of "anemic?" And does today's definition of "Computer Science" match up with that of 15 years ago? The same problem of definition can come into play when terms like "adequate curriculum support collection" or "research collection" are bandied about. One very real possibility is that a definition can be made so narrow, and so restricted to local needs, that the argumentation is circular, and all we prove by saying a collection is adequate is that we define our collection as adequate. No comparison is then possible with an external source of verification. This is one reason why a conspectus of library holding strengths, such as the proposed RLG/ALA conspectus, relies—just as surely as nuclear arms talks—on vigilance and outside verification.

Another assumption that we make when we read statistical data is that the degree of accuracy is credible. Sometimes we simply have to admit that precision is possible only within limits, and that certain facts are either unknowable or constantly changing. When the late sixteenth-century German physician named Weirus "computed" the number of demons on earth at precisely 7,405,926, divided into seventy-two battalions, we can only surmise whether or not he seriously believed his own figures.22 A more current example of spurious accuracy is when a cataloger informs us there are now 4,000,007 volumes in the library, since he or she has just cataloged 7 to add to the previously existing four million. Another example is the claim that "16.47% of the books present in United States university collections were published before 1887, compared to 23.96% of those in Great Britain." Rather than blindly accepting a level of precision that suggests an unimpeachable source, we do well to ask ourselves whether such accuracy is possible, given the thing being measured.23

Probably one of the most misused terms is that of the "average." The illogic of claims such as the following is obvious to all:

When I was a young librarian just starting out, I ordered a great number of books that should not have been ordered. As I gradually succumbed to budgetary constraints, however, I began not ordering books that should have been; so on the average, the collection came off all right."

Single observations can sound very dramatic when compared to averages that ignore the notion of dispersion; i.e., the fact of scattered numbers that may differ greatly from, but together produce, the average. It is meaningless nonsense to say, for example, that "the average American spends ten hours annually in prison." Even if that is the result of dividing total Americans into total prison hours spent, it ignores the fact that certain select citizens contribute much more than their share to that "average." It is just as confusing to assert that the "average book in our library is circulated for only twenty-seven minutes a year!" since it ignores the high percentage of books that are not circulated at all. If it were meaningful to find average time of circulation, then only circulating books would be included, and such differences as exist, e.g., between varying student and faculty loan periods, would also be taken into account.

After avoiding far-fetched estimates, doctored charts, puffed-up percentages, improper comparisons and faulty relationships,24 there remains the task of making sense of the factual information, reporting conclusions to interested parties, and guiding collection goals to reflect those findings.

Even if we librarians do have a sedentary job, two types of exercise we do not need are (a) skipping pertinent facts and (b) jumping to conclusions. Consider the following inductive leaps:

Fact A: "Twenty-two other equal-sized libraries in the nation have increased their overall collecting rates more than Library X since 1960."

Conclusion offered: The collecting efforts of Library X are inadequate.

Equally plausible alternative conclusion: The collecting rates of twenty-two other libraries in the nation may have been at abysmal levels before 1960, and they have had much more ground to make up than Library X, at least as regards relative rate of growth.
Fact B: “Less than half of the student body has even been into our library.”

Conclusion offered: We need a larger entryway.

Alternative conclusion offered: Our collection is not meeting student needs.

Equally plausible alternative conclusion: More than half of the students don’t need what we have to offer.

Fact C: “More journals are found mutilated in the current periodicals reading room than are found mutilated after students have returned them from homework assignments.”

Conclusion offered: We should store the current periodicals in students’ apartments.

Perhaps we would do well to scrutinize our own assessments as critically as we do the work of others. Although no simple rules apply, we can ask questions like: How reliable is the data source? Does anyone involved in the assessment have an “axe to grind”? What further supportive evidence is offered? Do the underlying assumptions seem reasonable? Do the estimates appear plausible? Does evidence point toward the support or rejection of initial assumptions? A good hypothesis is one that invites competing explanations and hypotheses.

Remember, a test of a statistical hypothesis can never prove anything; it merely adds evidence which either supports or discredits. We may want to ensure objectivity throughout the process of collecting, measuring, and analyzing data—and, indeed, in conceptualizing our problem at the beginning.

It has been said that when the truth or falsehood of an observation may have important bearings on conduct (in our case the conduct of library collecting), “over-doubt is more socially valuable than over-credulity.”

Faulty conclusions, grown on the vines of sloppy assumptions, can lead to bitter fruits. Conversely, a well-used statistical approach can enliven and clarify a collection assessment. May it never be said of us as librarians that we used statistics as a drunken man uses lampposts—for support rather than illumination.
Transfer of Materials from General Stacks to Special Collections

Samuel A. Streit

ABSTRACT. Recent concerns with collection development policies, preservation, and security have led libraries to consider transfer of materials from their general stacks to special collections as a part of collection management. The success of such a transfer program depends on the careful development of a policy, the establishment of clearly stated areas of responsibilities, the recognition of the importance of realistic procedures, and the setting forth of workable criteria for selecting material to be transferred.

To some degree, demonstrably rare materials have always been transferred from the general stacks of libraries into their special collections, but in the past few years this has become a trend, particularly in academic and public libraries that have substantial collections of older materials on their shelves.

While local considerations inevitably affect transfer decisions, there appear to be two or three general factors that in some combination have induced a rapidly-growing number of libraries to undertake a systematic program of sequestering portions of their holdings through transfer to special collections.

The first of these factors has to do with the collection assessment and collection development policies that many libraries have constructed as a way of coping with space limitations and diminished purchasing power. The process of devising collection development policies, while designed to build collections more rationally, has also served to focus attention on what the library already has. This assessment of retrospective holdings, in turn, has confirmed what has been known all along—that libraries harbor extensive collec-

Samuel A. Streit is Assistant University Librarian for Special Collections, Brown University Library, Box A, Providence, RI 02912.

Collection Management, Vol. 7(2), Summer 1985 © 1985 by The Haworth Press, Inc. All rights reserved.