

**STEEPED IN STICKY STATISTICS:
ADVENTURES IN COLLECTION DEVELOPMENT**

by

Jeannette Woodward

If recent surveys are accurate, we're each up to our eyelashes in personal computers. Hardly a library exists that doesn't have at least one Apple, IBM or clone on the premises. Consequently, librarians tout computer literacy almost as a new religion promising salvation from all library problems except bubble gum bookmarks.

If truth be known, however, most of those personal computers (PCs) are probably used for word processing and/or the sending out of overdue notices. There seems to be a common belief that until you have a super hi-tech integrated system with all the bells and whistles, more creative applications, like collection development, are out of the question. That may be because recent literature linking computers with collection development sounds so complicated. Possibly you think nothing of doing an analysis of variance on resource allocation on your trusty mainframe, loaded with a heavy duty statistical package. The thought of it fairly boggles the rest of us.

Is there such a thing as a middle-tech compromise? Are there some sensible ways of analyzing and strengthening the library's collection without amassing enough expertise to write a doctoral dissertation? Sure there are. Those little PCs can churn out mountains of data.

With the most primitive statistical methods you can achieve some surprisingly accurate insights.

In a nutshell, collection development means identifying areas which need to be strengthened because of high demand. In academic and school libraries, curriculum needs are another big consideration. In other words, are the books we're buying the ones we need? How can we tell? Even a non-math type can correlate book orders with circulation as long as records of some sort exist in both areas.

Librarians have a habit of filling the same holes several times over. Back when they discovered the library had nothing on AIDS or acid rain, they began stocking up, but like the sorcerer's apprentice, never knew when to stop.

So where can you begin? Let's start with your PC. It matters not whether it's an IBM or a Commodore. What does matter is how much memory you have to work with. Almost any collection development project you may undertake will require some sort of database management software. Naturally, the size of your memory limits the size of your database. There are ways of getting around this. Very large files can be manipulated with small memories if you want to put forth the effort, but why bother?

If your computer doesn't have at least 128k of RAM, consider spending the hundred dollars or so it costs to expand it. 256K of RAM would be even better and 512K or 640K ideal. The less you have to worry about memory size, the more attention you can devote to more interesting matters.

Next, you'll need a database program. If you already have one you like, go ahead and use it. Remember, however, that new programs can use information from other sources. Called importing data, this makes

it possible to use records that already exist, rather than keying them into your database. For example, information from circ, cataloging and acquisitions may already be available on disk. It would be a shame not to use it.

Now look on your library shelves for an introductory statistics text. Somewhere around chapter one, you should find an explanation of the correlation coefficient. This is a simple, though unsophisticated, statistic. As long as you remain aware that the numbers you are generating and the corresponding conclusions you read are distinctly suspect, you shouldn't go too far astray.

Why then, if conclusions are suspect, should you do this? The collection development or reference librarian at a college library may have the feeling that political science usage is concentrated in American government when all the new books seems to be about international relations, but there's not much she can do with impressions. If, however, she knows that the circulation of U.S. materials outnumber international materials by roughly five-to-one and the numbers are reversed when it comes to book orders, she has a good reason to confront the political science faculty. The statistician might find lots to quibble about, such as the amount of serendipity with which call numbers are assigned. Despite the eccentricities of the Library of Congress, the librarian has enough information to identify trends or tendencies though she lacks precision.

If you happen to have a statistical software program on hand, so much the better, but don't forget the textbook. If you don't know what the numbers you're getting mean, they're only so much gibberish.

If you're using an automated circ system, no matter how primitive, you have a big advantage. In fact, the usage reports such systems can print out are a very good reason to invest in one. Many libraries choose

to hold off using the smaller, truncated versions waiting for a Vax, Dynix or what-have-you necessary to run a more sophisticated system. If you know that moment is a piece down the road, you might want to look more closely at an interim system which can keep track of usage by call number. A year or two of complete (not sample) circ statistics arranged by exact call number is a treasure beyond price.

Before the system goes into operation, some libraries barcode only materials which have circulated at least once in the past five years. After that, barcoding is done as materials are checked out. Since the barcode number goes on the shelf list card, the percentage of barcoded cards within a particular call number range provides significant information about usage in the corresponding subject area.

That brings us to the subject of call numbers. Despite their limitations, the library statistician must depend upon them. They form the most reliable subject link between all the stray data floating about the library. As long as you have the call number, any two sorts of information (for instance book orders and circulation) can be correlated with each other. Librarians delight in regaling one another with the ridiculous number LC has assigned this or that book. Overall, however, the call number does indeed reflect the subject matter.

This is not to say that you can learn very much from the classification schedules. Before you begin, it will be necessary to translate all those numbers and headings like "special topics of general applicability" into something that makes sense in your particular library. Academic and school librarians should connect call numbers to curriculum emphases; the more specific, the better. Public librarians can identify popular topics or genres and their corresponding classification.

Now that you've got the machinery in place, you'll need some data. You have it lying around in all sorts of places. If someone on your

staff has a fetish about filing each and every copy of the MCOF in a different place, so much the better. However, you need the copy that comes out of cataloging, complete with call number. Next, it's important to get the best circulation data possible. An automated system makes it easier, but if none is available, at least get some good samples.

Circulation record books are available by LC letter or by hundreds if you're using Dewey. However, unless you break the categories down into smaller units, they're much too general to be of any value. Your data should be as specific as possible since even a decimal place can mean a totally different subject.

Your lost books file is another good source of data. You all know that sex, ESP, and witchcraft are the first titles to go. Beyond these topics, however, your losses pretty well reflect your usage. If you've been saving boxes and boxes of old catalog cards, now's the time to put them to use. Again, be sure you have a call number. Look around for other good sources of data like accessions lists from previous years, overdues, cataloging and reserve statistics.

No matter what sort of data you're using, the procedure is the same. Construct a database, input your data, arrange your output by call number and compare. If data is already available in machine readable format, you're all set. Nearly all circ systems can print out circulation records for a year or more by call number. Other records might need to be transferred to a more flexible database to manipulate fields efficiently. If so find out first in what formats the old database is capable of exporting information, then check the new one you wish to use to be sure it can handle the input. Many library application packages are essentially databases even though they may be called something else entirely.

If you feel more comfortable simply doing percentages (the number of titles in a particular classification range divided by the total number of titles circulated, ordered, etc.) that's fine, if a bit cumbersome. The correlation coefficient will take a bit more time to learn, but will be simpler in the end.

Interpretation is your next task and it's by far the most difficult. You may choose to start by looking at a particular subject area about which you already have some suspicions. In the political science example above, the librarian needed data to support a "gut" feeling. On the other hand, you might do better to approach your data more objectively. Look for particularly low correlations. They might surprise you. In school and academic libraries, there's often little connection between the faculty members who order most heavily and those who actually encourage library use among their students. However you approach the subject, be absolutely certain you get your money's worth out of all your efforts. Analyze your results and share them with everyone involved in selection.

With book costs as high as they are, it's vital to target your purchases to meet real and current needs. The statistical information you've gathered will enable you to put your limited resources to work where they're most needed.