

A Formula for Estimating Collection Use

A method is presented to estimate the proportion of books in a library which are responsible for the determination of the circulation performance rate of that library. The method is applied to a university library and to a public library with the finding that for the two libraries examined a much smaller proportion of the collection determines the circulation performance rate of the university library than for the public library.

IN A RECENT ARTICLE Daniel Gore defined the performance rate of a library as the measure of the percentage of all books patrons may want that are on the shelves when they want them. He suggested that a performance rate of 45 percent is typical for a university library.¹ In the same article he then proceeded to indicate how the work of Richard Trueswell can be used to determine a criterion for weeding books from a collection in such a way as to allow the performance rate to jump to 81 percent.² However, there is no theoretical foundation for the implications presented by Gore because the Trueswell criterion of last-circulation date measures *circulation* demand.

Current research has not firmly substantiated the idea that circulation use is a good indicator of overall use. Thus, Trueswell's technique actually affects the *circulation* performance rate, which is the probability that a book which a patron wants for circulation purposes is in the stacks when he or she wants it. Trueswell's technique trades a small percentage drop in the circulation performance rate for a large amount of

space in the stacks. Trueswell has, in fact, published results which indicate that only 60 percent of the books in the stacks are satisfying 95 percent of the circulation demand for books in the collection.³ In this case, 40 percent of the amount of space currently utilized would be traded for a reduction in the circulation performance rate of 5 percent of its current value. Gore's idea is that the performance rate could then be increased significantly by judiciously purchasing books to fill the amount of space created.

This article introduces a procedure which uses the last-circulation-date statistic to estimate the proportion of books in the stacks of a library which are responsible for the determination of the circulation performance rate. The estimate will not give us an indication of the rate value, but it will allow us to compute the percentage of books that actually are being used by the patrons for circulation. Results shown in this study indicate that this percentage may be much lower in university libraries than in public libraries. In fact, whereas Gore has implied that 50 percent of the cataloged books patrons want are not available on the shelves when they want them, this paper shows that at least 50 percent of those books

Stephen J. Turner is assistant professor, College of Business Administration, University of Akron, Akron, Ohio.

that *are* available for circulation are not used for circulation purposes. More will be said later after we have provided estimates of actual percentages.

One of the most appealing aspects of the estimation procedure is its simplicity. The librarian proceeds by conducting a random sampling of books from the stacks and circulation and then calculates the proportionate number of books having last-circulation dates in each sample. Our explanation of the estimation procedure requires the following definitions, and these terms will be used extensively in the description.

1. *Stack Population* (SP): A book is a member of the stack population if it is available to library patrons for outside use and is present in the library stacks. Suppose all books are removed from the stack population. Clearly, the circulation performance rate would be adversely affected to the extent that patrons who walked into the stack population area in anticipation of selecting a book for circulation would not find what they want. Equally as clear, if no books are removed, then there would be no effect on the circulation performance rate. Somewhere between *no* books and *all* books is a collection of books which could be removed from the stacks without affecting the circulation performance rate. If these books were removed from the stack population, then that collection of books remaining would be those which determine the circulation performance rate of the library. This leads us to the next definition.

2. *Circulation Core* (CC): A book is in the circulation core if it is a member of the special subsection of books in the stack population which determines the circulation performance rate of the library. A major proportion of the books which comprise the circulation core can be easily identified because they have a last-circulation date, as defined below.

3. *Last-Circulation Date* (LCD): The last-circulation date for a book in the stack population is its most recent due date. The last-circulation date for a book in circulation is its previous (as opposed to current) due date.

Perhaps the easiest way to show how the above classes of books are related is by means of the diagram shown in Figure 1, where:

SET	DESCRIPTION
U	Books that have been acquired by the library
CCB	Currently Circulating Books (horizontally lined area)
A	Books in the circulation core that have not previously circulated (dotted area)
SLCD	Books in the Stack population having Last-Circulation Dates (i.e., that have previously circulated)
B	Books in the stack population that are not in the circulation core (vertically lined area)
SP	Books in the Stack Population (union of sets A, SLCD, and B)

From Figure 1 note that the union of sets A and SLCD represents the set of books in the Circulation Core (CC).

Next let $N(CC)$ denote the number of elements in the set CC. Using this notation, the question of interest can now be stated as: How can the proportion

$$N(CC)/N(SP)$$

be estimated?

In nonmathematical terms, the above ratio is the proportion of books in the stack population which are responsible for the determination of the circulation performance rate. The calculation of this ratio can provide an indication of the number of books from which past circulation performance rates have been determined and could be an indicator of the number of books that could be weeded from the stack population with-

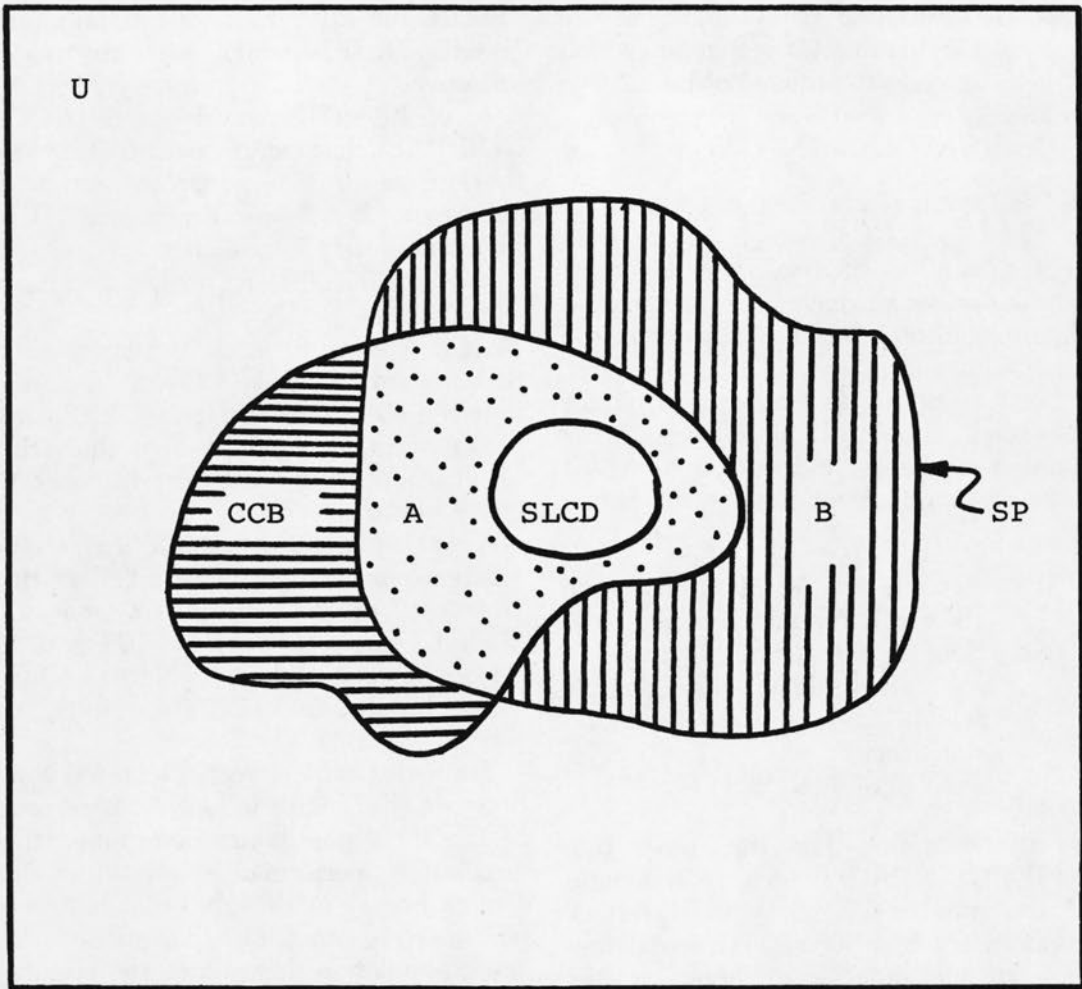


Fig. 1

The relationship between the books that currently are circulating (CCB), books in the circulation core (the union of sets A and SLCD), and books in the stack population (SP)

out affecting the rate for the future.

The last-circulation-date statistic can be used in computing an estimate for the above ratio, and a method of sampling the above sets can be developed which will allow the calculation of a confidence interval for the ratio.

First, we obtain a sample from the set of books in the Circulation Core by examining all books which are returned to the library from circulation during a given time interval. Second, we assume a scheme such that each of the returned books has (1) its checkout card stamped with two or more different due dates or

(2) has a checkout card that has been stamped with only one due date.

Our assumption is that each book will be a member of one of the two classes. In the first case the book is circulating for at least the second time and from the previous definition has a last-circulation date. In the second case the book is assumed to be circulating for the first time and does not have a last-circulation date.

Now let

CCS denote the set of books in the Circulation Core Sample;

CCSL denote the set of books in the Circulation Core Sample which have Last-circulation dates.

Then

$$N(\text{CCSL})/N(\text{CCS})$$

is an estimator of the proportion $N(\text{SLCD})/N(\text{CC})$, which is the ratio of the number of books in the stack population which have a last-circulation date to the total number of books in the circulation core.

Next, suppose a random sample of the stack population is taken and the number of books in the sample which have last-circulation dates is recorded. If

SPS denotes the set of all books in the Stack Population Sample and

SPSL denotes the set of all books in the Stack Population Sample which have a Last-circulation date, then

$$N(\text{SPSL})/N(\text{SPS})$$

is an estimator for the proportion $N(\text{SLCD})/N(\text{SP})$, which is the ratio of the number of books in the stack population which have a last-circulation date to the number of books in the stack population.

The two estimators above allow us to establish an estimator for $N(\text{CC})/N(\text{SP})$:

$$N(\text{SPSL})/N(\text{SPS})$$

is an estimator for $N(\text{SLCD})/N(\text{SP})$;

$$N(\text{CCSL})/N(\text{CCS})$$

is an estimator for $N(\text{SLCD})/N(\text{CC})$,

so that

$$r = \frac{N(\text{SPSL})/N(\text{SPS})}{N(\text{CCSL})/N(\text{CCS})}$$

is an estimator for $\frac{N(\text{SLCD})/N(\text{SP})}{N(\text{SLCD})/N(\text{CC})}$

$$= \frac{N(\text{CC})}{N(\text{SP})}$$

Thus, the ratio, r , is an estimator for $N(\text{CC})/N(\text{SP})$, which was our main objective.

A confidence interval for $N(\text{CC})/N(\text{SP})$ also has been developed,⁴ with the end result being that we can find values of LB (Lower Bound) and UB (Upper Bound) for which

$$\text{pr}(\text{LB} \leq N(\text{CC})/N(\text{SP}) \leq \text{UB}) \geq 0.95.$$

That is, we are *at least* 95 percent sure that the true ratio, $N(\text{CC})/N(\text{SP})$, lies between the values LB and UB. Examples now are provided which show the result of applying the theory to two different libraries.

A stack population sample and a circulation core sample were taken at the University of Massachusetts (Amherst) Main Library (UMass) in 1974 and at Forbes Public Library, Northampton, Massachusetts, in 1976. The results are summarized in Table 1.

From this table it can be stated that between 45.7 percent and 52.4 percent of the stack population determined the circulation performance rate for the UMass library in 1974 and that between 85.1 percent and 96.2 percent of the stack population determined the circulation performance rate for the Forbes library in 1976. We are *at least* 95 percent sure that the above bounds are correct.

CONCLUSIONS

The tabular results shown above indicate that there may be a significant difference between the proportionate sizes of the circulation cores in university and public libraries. Other samples from different libraries should be taken to see if this idea can be supported. The estimation procedure would key on the calculation of two ratios which require samples from the circulation core and stacks of the library. The details of the confidence interval calculation are omitted, but sample sizes of 3,000 books will normally be of sufficient size to guaran-

TABLE 1
 INTERVAL ESTIMATES OF THE PROPORTION $N(CC)/N(SP)$
 FOR THE UNIVERSITY OF MASSACHUSETTS AND FORBES LIBRARY COLLECTIONS

Sample	Collection	Number in Sample	LB	r	UB
stack population	UMass	2,286			
			.457	.490	.524
circulation core stack population	UMass Forbes	5,875 1,593			
			.851	.904	.962
circulation core	Forbes	1,155			

tee that the ratio, r , will approximate the theoretical ratio, $N(CC)/N(SP)$, to within ± 0.05 with at least 95 percent confidence.

The 0.49 value in Table 1 may be normal for a university library where large numbers of books are acquired each year and where the emphasis may be on the purchase of monographs which support institutional research objectives. This idea could be of interest from a management point of view if one desires to establish a rough estimate of a collection's circulation performance in relation to the same type of collection performance in other libraries

of similar composition and objectives.

REFERENCES

1. Daniel Gore, "The View from the Tower of Babel," *Library Journal* 100:1599-1604 (Sept. 15, 1975).
2. Richard W. Trueswell, "Analysis of Library User Circulation Requirements. Final Report." NSF Grant GNO435, January 1968.
3. Richard W. Trueswell, "Growing Libraries: Who Needs Them? A Statistical Basis for the No-Growth Collection," in Daniel Gore, ed., *Farewell to Alexandria* (Westport, Conn.: Greenwood Pr., 1976), p.72-104.
4. Stephen J. Turner, "The Identifier Method: An Analysis of the Last-Circulation-Date Approach for Measuring Library Collection Use" (Doctoral dissertation, Univ. of Massachusetts at Amherst, 1976).