DOCUMENT RESUME

ED 086 213 IR 000 062

AUTHOR Tolliver, Don L.; Lied, Terry R.
TITLE Determining a Cost Effective Sampling Technique Which Will Provide Estimates of the Number of Patrons Utilizing the Purdue General Library During the Fall Semester of 1973.
INSTITUTION Purdue Univ., Lafayette, Ind. Instructional Media Research Unit.
REPORT NO IMRU-09-73
PUB DATE Aug 73
NOTE 13p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Cost Effectiveness; *Costs; *Data Collection; Facility Utilization Research; Library Expenditures; *Library Surveys; *Sampling; Tables (Data); University Libraries; *Use Studies
IDENTIFIERS Purdue University

ABSTRACT

A study examined the feasibility of developing a cost effective sampling technique which would estimate the mean number of patrons using the Purdue University General Library during one semester. The technique employed, called random sampling without replacement, meant that, from the total population of days in the Fall Semester, particular days for counting would be chosen at random and used only once. Confidence bounds and cost figures were estimated for each sample size, based on the 1972 statistics of the library. Time-and-motion study techniques were employed in the re-examination of functions currently performed by the exit-checkers at the library. Check-out and exit procedures were modified to increase efficiency, and rough estimates computed of the daily cost of counting exiting patrons under the old and new procedures. In a university setting, library patronage is subject to such variables as vacations and exam times; therefore a stratified random sampling technique, where each day of the week was selected a given number of times, might yield lower error estimates than those computed by the "pure" random sampling method used in this study. (SL)

DETERMINING A COST EFFECTIVE SAMPLING TECHNIQUE WHICH WILL PROVIDE ESTIMATES OF THE NUMBER OF PATRONS UTILIZING THE PURDUE GENERAL LIBRARY DURING THE FALL SEMESTER OF 1973

By

Don L. Tolliver

and

Terry R. Lied

August 1973

IMRU-09-73

Introduction

Without question, current information and data on the dynamics of a large university library system is essential for its proper management and administration. However, increasingly, library administrators are faced with the need for more data to complete internal comparisons, to compare one library system with another, and/or to satisfy external requests for varied and more detailed data. These small but steady demands for additional data may cause once effectively collected data passed through efficient communication channels within the library system to degenerate.

In many instances, staff continue to spend significant amounts of time in data collection routines for which they may show little concern or have little knowledge of why such a task is undertaken. Complex and unwieldy data collection policies and procedures often become ingrained and accepted in library systems. Thus, those sensitive to the high costs incurred in such data collection routines are staggered by the efforts required to bring efficiency and logic to these data collection activities.

Predictably, library administrators reflect growing concern about the efficiency of such activities and the impact these traditional policies and procedures place on other pressing task priorities. Thus, they search for more efficient data collection techniques. Can a less expensive, yet reliable, procedure be established which will provide an adequate estimate of the number of patrons utilizing any given library for any given time period?

Objectives

The objective of this study is to develop a cost effective sampling technique which will provide estimates of the number of patrons utilizing

the Purdue General Library during the Fall Semester of 1973. Confidence bounds will be established for various sample sizes, and cost figures will be estimated based on functions performed by exit checkers.

Methodology

The current problem concerns whether or not a sampling technique can be developed that will yield an adequate estimate of the mean number of patrons who will utilize the Purdue General Library on a daily basis during the Fall Semester of 1973. In this case, a sample can be defined as that portion of the population which will be used to obtain an estimate of the mean number of patrons who daily utilize the General Library during the 1973 Fall Semester time period. For purposes of this study, the population from which the sample will be drawn will consist of the total days that the General Library is open during the Fall Semester of 1973. A sampling technique will be employed which Dixon and Massey (1969) refer to as random sampling without replacement. As applied to this problem, this technique is one in which the particular calendar days selected (from the total population of days available during the Fall Semester) for counting the number of patrons utilizing the library would be chosen at random and without replacement. Briefly, sampling without replacement means that an item (in this case, a particular day) could be chosen only once, e.g., if a particular calendar day has been selected and used by a random process, it cannot be selected and used again for it has not been placed back in the original population.

Let us assume that a large number of equal sized samples of Fall Semester days were drawn without replacement from the population of calendar days.

The number of calendar days in each of these samples is denoted by N. The distribution of the mean number of patrons utilizing the General Library in each of these samples has a standard deviation. This standard deviation is known as the standard error of the mean. The standard error of the mean is given by the following formula:

$$\text{S.D. of } \bar{x} = \sqrt{\frac{\sigma^2}{N}\left(\frac{N_p - N}{N_p - 1}\right)}$$

Where S.D. of $\bar{x}$ = standard error of the mean; in this case, the standard deviation of the sample distribution of means of the daily number of patrons who utilize the General Library.

σ^2 = the population variance; in this case, the variance of the daily number of patrons utilizing the library for one semester.

N_p = size of the population; in this case, total number of days the General Library is open during the Fall Semester of 1973.

N = size of sample; in this case, the total number of days chosen for sampling the number of patrons utilizing the General Library.

Since the standard deviations cannot be known beforehand, it is necessary to obtain this estimate by using a known population with similar parameters. An important assumption here is that the standard deviation of the daily mean number of patrons utilizing the General Library during the Fall Semester of 1973 is approximately equal to the standard deviation of the number of patrons utilizing the General Library on a daily basis during the Fall Semester of 1972. This assumption must be made for it is necessary to have an

estimate of the population standard deviation before error estimates can be computed. Thus, the standard deviation of the daily number of patrons utilizing the General Library during the Fall Semester 1972 was used to develop an estimate of the population standard deviation. It was found that the mean of the Fall Semester 1972 daily data was 1446 and the standard deviation was 724. It is possible that the size of the standard deviation for the Fall 1973 data could differ appreciably from the size of the standard deviation for the Fall 1972 data. However, as a check, the Fall 1971 standard deviation of patrons daily utilizing the General Library was computed, and it was found to be comparable with the Fall 1972 figure. Thus, this evidence supports the assumption of yearly comparability of standard deviations.

A review of the study of Purdue Libraries Statistics (IMRU-01-72) was made with particular attention focused on the section concerning the General Library, patrons leaving the library, pages A-10 through A-13. This was followed by a re-examination of functions performed by exit checkers both at the time of the original statistical study and, recently, under current conditions. Time-and-motion study techniques were employed in the re-examination of functions currently performed by these personnel.

The following methods were employed to determine a rough estimate of the daily costs incurred under both past and current conditions for counting the number of patrons exiting the library. Data reported in the 1972 statistical study were utilized to determine costs for past conditions.

Total annual seconds spent in counting patrons exiting the library	÷	Total number of patrons utilizing the library per year	=	Average time per transaction

Total number of patrons utilizing the library per year	÷	Mean number of transactions per hour	=	Total annual hours spent in counting patrons exiting the library
Total annual hours spent in counting patrons exiting the library	÷	Total days per year	=	Hours per day
Hours per day	x	Hourly rate	=	Daily cost

For current conditions, the methods were as follows:

Total number of patrons utilizing the library per year	÷	Mean number of transactions per hour	=	Total annual hours spent in counting patrons exiting the library
Total annual hours spent in counting patrons exiting the library	÷	Total days of the year	=	Hours per day
Hours per day	x	Hourly rate	=	Daily cost

Results

Table I presents the results in summary form. Also, this table illustrates the size of the error which might occur for various sample sizes selected to estimate the daily mean number of patrons utilizing the General Library during the Fall Semester of 1973. In making the computations in Table I, it was assumed that 110 total days constituted a semester. This assumption was based on the fact that the General Library operated 109 days during the Fall Semester 1972. Minor deviations from this figure would not significantly alter the values in Table I.

The first column (left to right) of Table I indicates the sample size or the number of days that counts might be made of the number of patrons utilizing the General Library. The second column is the error at the 68%

confidence level, and the third column is the error at the 95% confidence level. These figures estimate within plus or minus bounds the daily mean error that would be expected in the number of patrons utilizing the General Library for the Fall Semester of 1973 by using the sample size selected in the first column. For example, if a sample of 40 days were to be selected and the 95% confidence interval was used, one could determine the expected error in the number of patrons utilizing the General Library for the Fall Semester by the following means: Multiply the confidence interval found in Table I by the number of days in the semester. In this example, the figures would be ± 183 times 110 which equals ± 20,130. This number indicates that 95 times out of 100 the estimate of the total number of patrons exiting the General Library during the Fall Semester would be within ± 20,130 of the "true" total number of patrons exiting the General Library during that semester. Finally, to the extent that the Fall 1973 standard deviation is greater than the Fall 1972 standard deviation, the error will be greater than the Table I entry. Conversely, to the extent that the Fall 1973 standard deviation is less than the Fall 1972 standard deviation, the error will be less than the Table I entry.

The fourth column contains estimated cost figures for each sample size. These cost figures are based on the data reported in the July 1972 Study of Purdue Libraries Statistics (IMRU-01-72). Since the General Library now operates only two exits, it would be noted that the data for past procedures, e.g. column four, have not been interpolated to reflect costs for two exits rather than for four. In the original cost study, it was reported that in the General Library, 38 staff members spent a total of 3212.25 hours collecting statistics on the number of patrons utilizing the

library at an annual cost of $7,031.93 or roughly $1.57 for each hour the library was open. As illustrated below, these figures were broken down further to reflect average daily costs incurred for counting patrons exiting through four exits under past procedures.

Total annual seconds spent in counting patrons exiting the library		Total number of patrons utilizing the library per year		Average time per transaction
11,564,100	÷	1,250,000	=	9.25 seconds
Total number of patrons utilizing the library per year		Mean number of transactions per hour		Total annual hours spent in counting patrons exiting the library
1,250,000	÷	389	=	3213
Total annual hours spent in counting patrons exiting the library		Total days per year		Hours per day
3213	÷	365	=	8.8
Hours per day		Hourly rate		Daily cost
8.8	x	$2.00	=	$17.60

However, since the original cost study, there has been a significant change in number of exits and functions performed by personnel responsible for counting patrons leaving the library. Currently there are two exits, whereas, in the past there were four. Also, in the past, these personnel were primarily responsible for, in order of priority, (a) checking out books and other circulation functions, (b) checking briefcases, i.e., making sure all books that left the library were checked out, and (c) counting the number of patrons exiting the library. Currently these personnel are primarily responsible for (a) checking briefcases, i.e., determining that all books exiting the library have been checked out properly, and (b) counting the number of

patrons who exit the library. Also, the way they approach the task of counting has changed substantially. The counter is now held in the checker's hand on a nearly continuous basis as they check briefcases, etc. In the past, they would lay the counter down to check out books and handle other circulation duties and they they would have to retrieve the counter and count, or estimate a count, of the number of patrons who had just exited. It is quite likely that under past circumstances, the counting procedure was significantly more time consuming than it is under current procedures. Thus, as shown in column five, the recent time-and-motion study revealed that personnel counting patrons exiting the library were spending approximately one second per transaction to complete this task while previously they reported spending roughly 9.25 seconds per transaction. Thus, average daily costs under current procedures used to count patrons exiting the library break down as follows:

Total number of patrons utilizing the library per year		Mean number of transactions per hour		Total annual hours spent in counting patrons exiting the library
1,250,000	÷	3600	=	347
Total annual hours spent in counting patrons exiting the library		Total days of the year		Hours per day
347	÷	365	=	.95
Hours per day		Hourly rate		Daily cost
.95	x	$2.00	=	$1.90

TABLE I

EXPECTED ERROR IN PATRON COUNT WITH VARYING NUMBERS OF DAYS SAMPLED

Sample Size (No. of days in sample) N	Error (In no. of daily patrons) 68% Confidence Interval	Error (In no. of daily patrons) 95% Confidence Interval	Past Procedures Costs (Four Exits)	Current Procedures Costs (Two Exits)
1	± 724	± 1448	$ 17.60	$ 1.90
5	± 317	± 634	88.00	9.50
10	± 219	± 439	176.00	19.00
15	± 175	± 350	264.00	28.50
20	± 147	± 294	352.00	38.00
25	± 128	± 256	440.00	47.50
30	± 113	± 226	528.00	57.00
40	± 92	± 183	704.00	76.00
50	± 76	± 152	880.00	95.00
60	± 63	± 126	1056.00	114.00
70	± 52	± 105	1232.00	133.00
80	± 42	± 85	1408.00	152.00
90	± 33	± 65	1584.00	171.00
100	± 22	± 44	1760.00	190.00
110	± 0	± 0	1936.00	209.00

Conclusions and Recommendations

Several conclusions may be drawn from this study. Perhaps, the use of an alternate sampling method might reduce the size of the error estimates reported in Table I. One method would be a stratified random sampling technique which might specify that each day of the week be included an equal number of times in the sample. For example, if the sample size were 42 days, then each day of the week would be selected 6 times for inclusion in the sample. These 42 days would be selected from the total population of days in the semester. However, the use of this technique does not allow the utilization of a readily derived mathematical function to estimate the error. When no concise error estimates are required beforehand, the use of the stratified random sampling technique is a possibility.

Library use is a complex process involving a myriad of unidentified variables. Many events including patron utilization of a large university library do not occur on a purely random basis. There are periods of intense utilization followed by periods of little use of library facilities. Semester mid-terms, finals, weekends and holidays are examples of significant factors which must be accounted for in predicting total yearly use based on sampling techniques. Therefore, in the example presented in this paper, the stratified random sampling technique would likely yield lower error estimates than those found in Table I where "pure" random sampling was employed.

After studying the data, the following recommendations can be made:

1. A time-and-motion study of patron count procedures in departmental libraries needs to be made based on methods used to develop cost data for column 5 of Table I. This would provide a further look at the costs of data collection in the departmental libraries.

2. If costs still appear to be unusually large, then explore the advisability of employing stratified sampling techniques in the departmental libraries.

3. Unless the job function for the checkers varies significantly from current conditions, it is recommended that patrons be counted on a continuous basis as they exit the General Library.

References

Dixon, W. J. and F. J. Massey, Introduction to Statistical Analysis, Third Edition. New Y McGraw-Hill Book Company, Inc., 1969.

Drott, M. Carl, "Random Sampling: a Tool for Library Research" College and Research Libraries, March, 1969, Vol. 30, No. 2, pp. 119-125.

Tolliver, Don L., "Study of Purdue Libraries Statistics," IMRU-01-72, Instructional Media Research Unit, Purdue University Libraries and Audio-Visual Center, July 1972.

Winer, B. J., Statistical Principles in Experimental Design, Second Edition. New York: McGraw-Hill Book Company, Inc., 1971.