EDWIN W. REICHARD and THOMAS J. ORSAGH

Holdings and Expenditures of U. S. Academic Libraries: An Evaluative Technique

The current acquisitions expenditures and holdings of the nation's academic libraries as a whole are examined relative to the numbers of students and faculty for the period 1952-1962 by the use of random sampling and multivariate analysis. The formulae which are derived from the data serve two functions: they describe existing national behavioral patterns; and they permit one to measure his institution's performance against that of other, similar institutions. It should be noted that the evaluative technique developed measures individual library performance against observed behavior rather than against a predetermined arbitrary standard.

IT IS BECOMING increasingly evident that librarians and academic administrators must seek objective performance standards to substantiate, in part at least, the need for the budgetary increases which will be required if the library is to remain a viable part of the educational program. Librarians must in any event make recommendations concerning such increases. These recommendations are usually expressed numerically and implicitly convey an aura of precision. They may take the form of a given percentage of the total institutional budget, a dollar amount of expenditure per student, a minimum gross number of volumes to be acquired within a certain time period, or perhaps a given number of volumes per student.

In general, it may be stated that a librarian's recommendations derive from two basic sources:

1. an intimate knowledge of the short-term educational changes at his institution, *e.g.*, the planned institution of a new degree program which will necessitate acquisitions in an area not heretofore well developed. Such recommendations are clearly *ad hoc,* are not easily analyzed, and therefore will not be considered in this study;

2. the application of some standard of long term growth. While it is theoretically possible to establish a standard or bench mark of growth *in vacuo,* in practice one usually compares one's own collection to those of other institutions, or to some arbitrary standard established by an independent agency such as ALA.

In order that any such comparison be acceptable and convincing to an administrator, it would appear requisite that

Mr. Reichard is Assistant Librarian for Readers' Services, and Dr. Orsagh is Assistant Professor of Economics, in Lehigh University, Bethlehem, Pennsylvania.

the criteria of comparison be unambiguous, quantifiable, and reasonable. In other words, one should relate his performance to those institutions which are comparable in clear-cut definable ways, rather than to those institutions with which one may subjectively *like* to compare himself.

It is the purpose of this study to provide the policy-maker with such quantifiable, unambiguous means of comparing his own performance with other, similar institutions and, as a byproduct of this effort, to provide information concerning the performance of the nation's academic libraries taken *in toto*, which may be of value to present and future standard-makers.

Part One

Before entering into a discussion of the techniques to be employed in the study, it is necessary to select the following:

1. the criteria by which we shall consider one institution comparable to another; and
2. the criteria for comparative library performance.

The criteria for (1) appear to present the greatest difficulties. Such things as educational philosophy, quality, and orientation are extremely difficult to measure. We have chosen instead three variables which have the virtue of being easily measured and which also are frequently considered to be relevant for comparative purposes:

1. The size of the undergraduate body is frequently the first relationship considered when the question of library size is raised. It is interesting to note that the value of this type of relationship has been questioned by Ellsworth.[1]

2. It has been a truism in the library profession that the great universities with the great faculties are the ones with the great libraries. It must follow, then, that the size of the faculty (the one quantifiable aspect of a faculty) be included in this analysis.

3. It can be said, from the point of view of the library, that graduate students are more nearly like faculty than are undergraduates in terms of their need for research material. In addition, within the past decade graduate programs have often been the source of the greatest growth within the universities; hence, graduate students shall be considered as a separate category in this study.

Since the past decade has seen a substantial and unequal growth in these variables, it was decided to observe them over an interval of approximately ten years. While the selection of these particular variables may seem arbitrary, they do give us an unambiguous and, we feel, valid source of comparison of institutions. The statistical results to be provided shortly will amply confirm this opinion.

With respect to the library performance criteria, we have selected only two: (1) holdings; and (2) expenditures for current acquisitions. These are entirely quantifiable and seem to be commonly cited. It might be noted that current acquisitions measured in volumes might have been used instead of current expenditures, but such data are not readily available and for practical purposes, expenditures will serve equally well. We have eliminated other current and capital expenditures because of their high degree of variability, which proscribes any useful comparative analysis. Consider, for example, two institutions with the same holdings and acquisitions rate: one may be decentralized and/or more heavily staffed in the public service division; or it may pay for such services as maintenance from the current administrative budget. Again, it should be pointed out that no attempt has been made to evaluate such subjective factors as the appropriateness of the collection

[1] Ralph E. Ellsworth, "The Legislature Is Not Convinced," *Library Journal*, XC (May 1965), 2199-2203.

to the institution or its value to the library user.

Part Two

Now that the criteria of performance and comparison have been selected, the nature of their interrelationship must be determined. On the surface it would seem as though a simple linear correlation would suffice; but, as we will show by means of an example, this technique can lead to erroneous conclusions because it ignores the complexities of these interrelationships. (The example will also introduce those readers unfamiliar with correlation technique to the concepts basic to the more sophisticated relations which will be established.) Consider the relation between a library's expenditures for books and the number of its undergraduates. To measure the degree of relationship between these two variables, a sample of academic libraries was drawn for 1952 and for 1962.[2] The sample excludes two-year colleges, vocational schools, extension schools of state universities, and other similar non-degree-granting institutions. The expenditures and the number of undergraduates associated with these libraries was obtained, and a correlation was run between the two variables. The resulting coefficients of determination for 1952 and 1962 were, respectively, $r^2_{52} = 0.53$ and $r^2_{62} = 0.48$, which means that 53 per cent of the variation, or differences, in expenditures by these libraries in 1952 and 48 per cent in 1962 can be explained by variation in the number of undergraduates.[3] These values are quite large, which means that we can be almost certain that if we had examined all twelve hundred libraries at each time period, we would have found a reasonably high coeffcient of determination for each year.[4]

Since r^2 is rather large, we would seem to have support for the hypothesis that libraries regard the size of the undergraduate student body as a criterion for determining the size of their book budgets; or possibly the converse hypothesis, *viz.*, that the size of the budget is a determinant of the size of the undergraduate student body; or, perhaps still better, that these two factors interact, producing the observed relationship. Notwithstanding the existence of this strong relationship, we are not required to accept any of these hypotheses, however. The relationship could be, and indeed in this case is, a spurious one. To indicate why this is so, and to illustrate the difficulty involved in the use of two-variable correlations, it is necessary to consider another relation, that between expenditures and the number of faculty. For 1952 and 1962 the coefficients of determination are, respectively, 0.71 and 0.80. This means that 71 per cent of the variation in expenditures in 1952 and 80 per cent in 1962 can be explained by

[2] A simple, statistically random sample of approximately three hundred institutions was drawn for each of the years. The population consisted of approximately twelve hundred such institutions. For these and later computations, part-time persons (undergraduates, graduates, and faculty) are given a weight of one-half. The data used throughout this study are taken from *The American Library Directory* (20th and 24th editions; 1954 and 1964); and *American Universities and Colleges* (6th and 9th editions; 1952 and 1964). It should be noted that the data are not reported for exactly the same years. Student enrollment and faculty data pertain to the Fall of 1951 and 1962; holdings and expenditure data, mostly to 1952-53 and 1962-63.

[3] r^2 can be defined in two meaningful ways: (1) it is a measure of the degree of association between two variables; its minimum and maximum possible values are, respectively, zero and one; (2) r^2 is a ratio, the denominator of which is the measure of average error associated with predicting, without "assistance," the value of a particular dependent variable (expenditures), and the numerator of which is the amount by which this error is reduced when one has the assistance of information concerning the relationship of the dependent variable to some independent variable (such as number of undergraduates). Thus, $100r^2$ measures the percentage reduction in the average error of prediction associated with the introduction of an independent variable to help in predicting the value for the dependent variable. The range of r^2 is, again, zero to one, since the minimum reduction in errors of prediction is zero and the maximum reduction cannot exceed 100 per cent.

[4] Precisely, there is less than one chance in forty that r^2_{52} is less than 0.45 or that r^2_{62} is less than 0.40 for all twelve hundred academic libraries.

variation in the size of faculty. These coefficients, too, are quite large[5] and would seem to lend support to one of the three hypotheses cited above, adapted, of course, to the new independent variable, size of faculty. Which of the hypotheses are correct?

Some light is shed on the problem when a third relation is considered, that between the two so-called explanatory variables, number of undergraduates and faculty. The evidence shows that these two variables are, themselves, closely related; normally, the more undergraduates there are, the more faculty there are. (The coefficients of determination for the two years are, respectively, 0.70 and 0.67.[6] Since expenditures are closely related to the number of faculty, and since the latter is closely related to the number of undergraduates, we may well get a close relationship between expenditures and undergraduates which is purely the effect of the number of faculty. More concretely, though probably less correctly, a library's expenditures as well as the number of undergraduates may be determined by the number of faculty.[7]

Clearly, two-variable correlations would be inadequate, or rather, misleading descriptors for our purpose. The obvious alternative, and the one we propose to adopt, is to use a multivariate analysis. This proposed procedure will allow us to isolate the separate effects of each of the independent variables and, incidentally, to incorporate into the discussion, simultaneously, the effects of the third independent variable, the number of graduate students.

[5] There is less than one chance in forty that r^2 for 1952 and for 1962 would be less than 0.65 and 0.75, respectively, for all twelve hundred libraries in the population.

[6] There is less than one chance in forty that the real coefficient is less than 0.64 for 1962 and less than 0.60 for 1962.

[7] An oftquoted analogy can stand service here: There is a very close relationship between the number of storks nests found in various sections of northwestern Europe and the human birth rates in those same sections—the more nests, the more human births. Mamma may have been right, but there is one other explanation of merit, *viz.*, that the number of storks nests and the number of human births are both correlated with the number of buildings present in an area, and that this latter variable (number of buildings), in fact, is the determinant of the other two.

Part Three

Now that we have shown the necessity for using multivariate analysis, we can proceed to develop our interrelationships. The relevant statistical data, obtained from four random samples, yield the following multiple regression equations:[8]

$$V_{52} = 51{,}700 - 105U - 37G + 1640F$$
$$(11) \quad (23) \quad (100)$$
$$s = 276{,}700$$
$$R^2 = 0.71$$

$$V_{62} = 27{,}100 - 9.6U - 59G + 969F$$
$$(5.0) \quad (19) \quad (63)$$
$$s = 147{,}600$$
$$R^2 = 0.75$$

$$E_{52} = 847 - .07U - .04G + 115F$$
$$(.76) \quad (1.3) \quad (11)$$
$$s = 20{,}700$$
$$R^2 = 0.71$$

$$E_{62} = 5910 - 4.7U + 39G + 279F$$
$$(1.7) \quad (7.3) \quad (21)$$
$$s = 57{,}000$$
$$R^2 = 0.82$$

V, E, U, G, and F refer, respectively, to number of volumes, dollars of expenditure for current acquisitions, and number of undergraduates, graduates, and faculty. The subscripts indicate the years to which the equations apply.

R^2, the coefficient of multiple determination, has essentially the same meaning as its two-variable counterpart, r^2; it measures the reduction in errors of prediction of V and E which result from using the three independent variables, U, G, and F. For example, the average error associated with predicting the size

[8] The random samples, of size three hundred each, were derived from the sources cited earlier.

of the library collection in 1952 is reduced 71 per cent by virtue of our using the first regression equation as the predictor. The fact that R^2 is less than one indicates that there are other variables which determine, or are related to, the number of volumes; but since R^2 is as high as it is, one can be reasonably confident that he has found the three variables which explain the largest proportion of interlibrary variation in number of volumes held.[9] These three variables taken together, therefore, may be regarded *for the purposes of this study* as the criterion to be used for the determination of the number of volumes held by United States academic libraries in 1952. Parenthetically, it should be noted that no cause and effect relationship is implied or intended: the three variables, U, G, and F may have determined V; V may have determined them; or V, U, G, and F may have been mutually determining.

The other coefficients of determination given above are at least as large as the first one; hence, all four of these multiple, linear regressions are satisfactory predictors. These regression equations, themselves, warrant examination. The numerical values attached to the U, G, and F symbols (the slope coefficients) are of particular interest. Generally speaking, the value of the slope coefficient is a measure of the average increase in the dependent variable, V or E, associated with a one unit increase in the value of the associated independent variable, U, G, or F, *assuming the other two independent variables do not change*. For example, conceive of two libraries whose undergraduate and graduate student bodies were the same in 1952 and whose faculties differed in size by one person. The one having the larger faculty on the average had 1,640 more volumes and spent $115 more on books.

The values in parentheses beneath the slope coefficients, the so-called standard errors of the slope coefficients, are also of interest. If one adds and subtracts twice the standard error from the value of the slope coefficient, he obtains a range of values which very likely will contain the true slope coefficient value; *i.e.*, the value which would, in fact, have been obtained had the regression equation been derived from data for all United States academic libraries. For example, one can be fairly sure that the increase in expenditures associated with a one-person increase in faculty in 1952 was between $93 and $137.[10]

So much for the meaning of the equations and their associated values.[11] What of significance do these statistics tell us? The strikingly important fact to be derived from them is that the size of faculty was the overwhelmingly important variable associated with both the size of the collection and the level of expenditures of these libraries. The singular lack of importance, or rather negative influence, of the undergraduate is also of considerable interest. In 1952 those libraries whose graduate and faculty personnel were of similar size, but whose undergraduate student size was the larger, had between 83 and 127 *fewer* volumes per undergraduate student. For number of volumes in 1962 and for expenditures in both time periods, the value was close to zero, indicating for these variables that the undergraduate's influence here was probably negligible. The relationship of V and E to G is of a mixed character. Those libraries having larger graduate enrollments, *ceteris paribus*, may have had fewer volumes in 1952—though, the large standard error makes this uncertain—and very likely had fewer volumes in 1962. In 1952

[9] There is less than one chance in forty that R^2 for all twelve hundred libraries would be less than 0.67.

[10] More precisely, there is approximately a 95 per cent chance that the following statement is correct: With respect to all United States academic libraries, the average increase in E in 1952 associated with a one-person increase in faculty was between $93 and $137.

[11] The s values, the residual standard deviations, will be brought into use in Part Four.

there seemed to be no relationship between graduates and expenditures, but in 1962 there was a decided, though small, positive relationship.

The four equations under discussion thus describe the typical behavior of libraries in 1952 and 1962; *i.e.*, they show the state of the world in these two time periods. But what of the changes which occurred between these two times? What do the equations tell us? Consider the equations dealing with number of volumes. Between 1952 and 1962 the coefficient of F declined from 1640 to 969. This means that on the average 671 fewer volumes were added to the collection per unit increase in faculty in 1962 than were added in 1952. The coefficient of G also declined, but much less than that of F; and that of U actually increased. Thus, one can say that the faculty of 1962 had a much smaller effect on the library collection of 1962, that the graduate student had a slightly more negative effect, and that the undergraduate had a definitely less negative effect. So much is true. It is important, however, that one *not* infer from these changes in coefficients that the typical library of 1952 was transformed into the typical library of 1962, and that the changes in these three coefficients reflect that transformation. One cannot draw this inference because the 1952 and the 1962 regressions are based on different populations.[12] The coefficients for 1952 (1962) are average values which describe the performance characteristics of the typical library of the population of 1952 (1962). Hence, the changes in coefficients will reflect both changes in the performance characteristics of the typical library *and* changes in the library population itself. In general, an average value, such as a regression coefficient, can change because of changes in the values of the elements making up the average or because of a change in the number of elements contributing to the average. The latter represents a change in the composition of the population. To take a specific example, suppose libraries associated with small graduate schools have lower-valued G coefficients than libraries with large graduate schools; and suppose over time the majority of libraries that come into existence have small graduate schools. Then, even if the G coefficients of each and every library, large and small, stay constant, the average value—the one generated by our sample—will decline in going from the 1952 regression to the 1962 regression.[13] As the ice melts, one's scotch gets weaker, but the alcohol content of the two ingredients, taken separately, does not change. Here it is seen that the decrease in the G coefficient arises solely from the increase in small graduate programs. The individual classes of libraries experience no change in coefficient—*i.e.*, the –5 and –69 values apply to both 1952 and 1962.

Size of Graduate Program	1952		1962	
	No. of Schools	G Coefficient	No. of Schools	G Coefficient
Large	10	– 5	10	– 5
Small	10	–69	50	–69

Weighted Mean Value of G (where G represents the sum of the coefficient times its own number of schools, divided by the totaled number of schools):

$$1952: -37 = [(-5)(10) \times (-69)(10)] \div [10 \times 10]$$
$$1962: -59 = [(-5)(10) \times (-69)(50)] \div [10 \times 50]$$

For some purposes, such as comparing one library's response to changing enrollment and faculty size with the response of other libraries over the course

[12] Some of the changes in population were quite dramatic. For example, the mean number of volumes per library, based on our sample data, increased by 24,149; while interlibrary variability (measured by the standard deviation of number of volumes) almost halved. Mean per annum expenditure per library increased by $46,025, while interlibrary variability in expenditure increased about 350 per cent.

[13] A numerical example may be of value. The following table contains one of the many possible sets of G coefficients which are consistent with the –37 and the –59 values generated by the regression equations for number of volumes. (Of course, we do not know what the values of the coefficients for large and small libraries actually are, but certainly the values given here are not unreasonable.)

of time, it is necessary to eliminate the influence of changes in the composition of the library population.[14] The obvious and conventional method for neutralizing population changes is to hold the population constant while drawing the sample; that is, one would examine the same set of libraries in both time periods. To this end, a sample of the libraries which were in existence in 1952 was selected, and the corresponding V, E, U, G, and F values were then obtained for 1952 and for 1962.[15] From these two sets of data a new set was created by subtracting each 1952 observation from its corresponding 1962 observation. This new set, consisting of the changes in V, E, U, G, and F, produced the following two regression equations:

$$V' = 27{,}200 - \underset{(8.4)}{16U'} + \underset{(26)}{125G'} + \underset{(72)}{629F'}$$

$$s = 132{,}700$$
$$R^2 = 0.55$$

$$E' = 14{,}800 - \underset{(4.2)}{3.5U'} + \underset{(13)}{93G'} + \underset{(36)}{270F'}$$

$$s = 67{,}310$$
$$R^2 = 0.58$$

The symbols have essentially the same meaning as before. The prime indicates that we are considering the changes in V, E, etc., between 1952 and 1962. How does one interpret the slope coefficients? Consider the coefficient of F′ for volumes. Imagine two libraries with undergraduate and graduate enrollments that increased by the same amount between 1952 and 1962; and suppose the faculty of the one increased somewhat more than the other. The library whose faculty increased more acquired approximately 630 more volumes per extra faculty person. In other words, the school whose faculty grew more rapidly, *ceteris paribus*, had increased its book collection by 630 volumes per added faculty member between 1952 and 1962.

One notes that the changes in faculty size dominate the changes in volumes and expenditures, that the changes in undergraduate enrollment, abstracting from changes in the other two independent variables, exerted little influence. Of some interest is the enhanced status of the graduate student. Those schools whose graduate programs were expanding were adding substantially to their collections and also to their book-purchasing budgets. The positive G′ coefficient for volumes implies that libraries were responding to the general expansion of graduate schools, and were adding, per student, approximately 125 volumes to their collections. But this raises a question: How can it be that the G (not G′) coefficient was significantly lower in 1962 (−59) than it was in 1952 (−37)? One can only conclude that the positive G′ tendency was offset by the initiation and expansion of graduate programs on the part of schools which, on the average, possessed small collections.

Part Four

The indexes of national performance which have been provided by our six regressions have direct relevance to one's own library. These regressions permit one to compare his own library's performance to that of other, similar libraries. The regressions also permit one to determine the level of expenditure or the size of the book collection which would be required if his library is to attain a particularly desired ranking among libraries of its own class. The procedure for the first case is quite simple. For the particular V, E, V′, or E′ comparison that one is considering, he computes a statistic, t^*, as follows:

[14] If this is not done, one might misinterpret the statistical results. The first example in Part Four will make this clear.

[15] Each of the regressions is based on a statistically random sample of size 150. One could have based the samples on the 1962 library population. The selection of 1952 was arbitrary but presumably not important.

$$t^{\circ} = \frac{y - y^{\circ}}{s},$$

where y is the value associated with one's own library; y° is given from the regression equation; and s is obtained from the collection of statistics which are attached to the regression equations. After deriving the t° value, one consults a special table, usually referred to as the table of the t probability distribution, which can be found in any ordinary statistics textbook. For the reader's convenience, an extensive summary of the relevant portion of that table is given below.[16] One consults the table to determine where his t° value falls within the table's array of t values. The corresponding percentile value then tells one where he is located with respect to other, similarly circumstanced libraries.

For convenience, let us refer to our own institution as Mythical U. Table 2 contains the essential statistical data for our university.

[16] One would enter a really extensive t table at 300 degrees of freedom for V and E values, and at 150 degrees of freedom for V′ and E′ values. (Usually some interpolation is required.) The values in Table 1 are an average of the two; but since the differences are quite small, no appreciable error can arise from using these approximations.

TABLE 1.

PERCENTILES OF THE *t* PROBABILITY DISTRIBUTION

t	Percentile Corresponding to *t*
2.61	99.5
2.35	99.0
1.98	97.5
1.65	95.0
1.29	90.0
0.84	80.0
0.52	70.0
0.25	60.0
0.00	50.0
−0.25	40.0
−0.52	30.0
−0.84	20.0
−1.29	10.0
−1.65	5.0
−1.98	2.5
−2.35	1.0
−2.61	0.5

The procedure for computing Mythical U's standing is laid out in Table 3. Column 2's values are obtained directly from Table 2; column 4's from the regression equation data which were presented earlier. Column 3's values are computed from the regression equations themselves. For example, for V_{32},

$$y^{\circ} = 51{,}660 - 105(3100) - 37(470) + 1640(310) = 217{,}000$$

and for E′,

$$y^{\circ} = 14{,}800 - 3.5(-50) + 93(230 + 270(30) = 44{,}460.$$

Calculating t° is then straightforward. For example, for V_{52},

$$t^{\circ} = \frac{390{,}000 - 217{,}000}{276{,}700} = +0.62,$$

and for E′,

$$t^{\circ} = \frac{55{,}100 - 44{,}460}{67{,}310} = +0.16.$$

The percentiles (column 6) are then obtained by entering Table 1 with the appropriate t° value. For example, $t^{\circ} = 0.62$ falls between the t values, 0.52 and 0.84, but is closer to the former; *i.e.*, our t° value is closer to the 70th than to the 80th percentile.

What, then, do we learn about Mythical U's library? We discover that in 1952 only about 30 per cent of the libraries which were in the same class as Mythical U—*i.e.*, possessing approximately the same numbers of undergraduates, graduates, and faculty—had a book collection which was larger than Mythical U's, and that only about 5 per cent of the libraries in Mythical U's class in 1962 had a collection which was larger than Mythical U's collection at that time. On the other hand, the university's expenditure level did not occupy as high a ranking as its collection. Moreover, its ranking declined drastically, and in 1962 was well below the average for its class. What are we to infer from these seem-

TABLE 2.

STATISTICAL DATA FOR A HYPOTHETICAL UNIVERSITY

VARIABLE	YEAR			
	1952		1962	
Volumes	....	390,000	....	524,000
Expenditures	....	$ 40,000	....	$ 95,100
Undergraduates:				
Full-time	2,800	.....	2,900	.....
Part-time	600	.....	300	.....
Total°	....	3,100	....	3,050
Graduates:				
Full-time	170	.....	200	.....
Part-time	600	.....	1,000	.....
Total°	....	470	....	700
Faculty:				
Full-time	290	.....	290	.....
Part-time	40	.....	100	.....
Total°	....	310	....	340

° Part-time counted at one-half of its actual value.

TABLE 3.

SAMPLE LAYOUT FOR COMPUTING PERCENTILES

Variable	y	y°	s	t°	Percentile
V_{52}	390,000	217,000	276,700	+0.62	Over 70
V_{62}	524,000	286,100	147,600	+1.61	Under 95
E_{52}	$ 40,000	$ 36,260	$ 20,700	+0.18	Under 60
E_{62}	$ 95,100	$113,500	$ 57,000	−0.32	Under 40
V′	+ 134,000	+ 75,580	132,700	+0.44	Under 70
E′	+$ 55,100	+$ 44,460	$ 67,310	+0.16	Under 60

ingly contradictory statistics? One very reasonable hypothesis is that the 95th percentile value arose from shifts in the characteristics of the libraries belonging to Mythical U's class—specifically to the presence of a larger proportion of schools with small collections—and that the 40th percentile value arose from these schools with small collections trying to catch up. Thus, the improvement in the one rank and the deterioration in the other in no wise reflect either favorably or unfavorably upon the university. Of course, if the new entrants to Mythical U's class have higher standards, on the average, than the earlier set of libraries had, then ultimately the university's position both in terms of expenditures and in terms of its book collection will be less favorable than it was in 1952. Now suppose that one abstracts from these changes in class composition; what then can be said of our university? We note that the additions which Mythical U made to its collection, V′, ranked slightly under the percentile value for V in 1952. The same is true of E′. We also note that the rank of E for 1952 was much less than that of V for the same year. These statistics show that Mythical U was not adding to its collection in sufficient amounts to maintain its 1952 rank for library collection. Even if the class composition had not changed, Mythical U would most likely have held a lower rank in respect to its collection in 1962.

We now turn to the second use to which the regression analysis may be put, that of determining the particular expenditure or acquisitions level which is consistent with a predetermined percentile ranking. A numerical example will illustrate the technique to be employed. Suppose that the 1962 Mythical U decided to embark upon an expansion program, and suppose that it wished to maintain an exact 95th percentile ranking with respect to its library collection, how much would its collection have to grow so as to maintain this rank? Suppose plans call for an expansion in undergraduate enrollment of five hundred, in graduate enrollment of six hundred, and in faculty of eighty-six. (The latter increase would maintain approximately the same student-faculty ratio.) These values allow us to generate the number of volumes held by the typical university of this new size.

$$V = 27{,}100 - 9.6(3550) - 59(1{,}300) + 969(426) = 329{,}400$$

A 95th percentile value implies $t = 1.65$. We solve the t^* equation for y.

$$t^* = 1.65 = \frac{y - 329{,}400}{147{,}600}$$

whence, $y = 572{,}900$. Thus, if Mythical U is to be ranked in the 95th percentile, it will have to add 572,900 – 524,000, or 48,900 volumes to its collection. Of course, the reliability of this estimate depends upon the extent to which the national performance standard does not change; *i.e.,* the extent to which the 1962 regression equation continues to be valid.

Conclusions

The statistical data presented in this study have contained some surprises. One might not have expected the undergraduate to have been as unimportant as he turns out to be, nor that the faculty would be so overwhelmingly important. The "oughts" and "shoulds" uttered by library administrators take on new meaning and may well require respecification, given that we now know something of what academic libraries in this country are and have been doing. Of course, the approach outlined in this study does not constitute a full solution to the administrator's problems. It only gives him the means with which to formulate those "shoulds" and "oughts" which involve measurable comparisons with other institutions. For the wide range of problems that do not admit of interinstitutional comparisons, the librarian will still have to search his own soul.

At a higher level of consideration, it is worth observing that academic libraries are highly predictable institutions —at least as far as the size of their collection and of their expenditures for books is concerned. With just two variables, the number of graduate students and the number of faculty, we can explain more than two-thirds of the interlibrary variation in volumes and expenditures. Hence, it is not unreasonable to suppose that a still better predictive equation can be developed by the introduction of new variables, with or without a reformulation of the definition of the existing two significant variables.[17]

In general, one achieves greater predictability when the objects of measurement are consistently and precisely defined. We recognize that the raw data used in this study have shortcomings by scientific standards. The size of R^2, however, strongly indicates that, despite their deficiencies, the data were adequate for our purposes. ■■

[17] Technically speaking, one can always increase R^2 by adding more independent variables. With a sample of size n, n–1 independent variables always yield $R^2 = 1.0$. One stops adding variables when the increase in R^2 is no longer statistically significant. Beyond this point, the increase in R^2 is regarded as trivial.