KENDON STUBBS

University Libraries: Standards and Statistics

The ARL-ACRL Standards for University Libraries *do not present quantitative standards, but rather place their emphasis on the performance of university libraries. Through the statistical techniques of correlation and regression, discriminant analysis, and principal component analysis it is possible to analyze university library data and to derive minimal criteria that statistically distinguish university libraries from other kinds of academic libraries. These criteria look very much like standards, but still fail to relate library size and resources deployed to library performance.*

THE ARL-ACRL *Standards for University Libraries* resolutely eschew numbers.[1] How many books does a university library need? The *Standards* reply: "A university library's collections shall be of sufficient size and scope to support the university's total instructional needs and to facilitate the university's research programs." How many staff members? "A university library shall have a sufficient number and variety of personnel to develop, organize, and maintain such collections and to provide such reference and information services as will meet the university's needs." How large a budget? "Budgetary support for the university library shall be sufficient to enable it to fulfill its obligations and responsibilities as identified in the preceding standards." There is a kind of sameness of sufficiencies here, which may seem fuzzy to those who want to know whether a particular library has an adequate budget or enough staff. One is tempted to regard the *Standards*, in Hegel's phrase, as the night in which all cows are black. The *Standards for College Libraries* by contrast appear almost blatant in quantification: A college library should have 85,000 volumes, plus 100 volumes for each FTE faculty member, 15 volumes for each FTE student, etc.; one librarian for each 500 FTE students up to 10,000, one for each 1,000 students above 10,000, etc.; .10 square feet per volume for the first 150,000 volumes, etc.; and so on.[2]

Nevertheless, the *Standards for University Libraries* offer an argument particularly attractive for these days: that a university library should be judged not by its size in collections or expenditures or staffing but by how well it serves students, faculty, and other academic staff. In fact, unlike the college standards, the *University Standards* begin with a section on services rather than collections. Whether a student can find the information he needs when he needs it is a more important test of a library, the *Standards* are saying, than whether the library has attained the more or less artificial goal of some minimum number of volumes. In a way it is this emphasis on services that hinders or precludes the formulation of quantitative standards. Up to the present, library data on system responses to user needs have not been adequate for establishing acceptable quantitative standards. In the remainder of this paper, as we derive what may look like quantitative standards, keep in mind that it is the *Standards for University Libraries*, in their emphasis on services and performance, that are putting first things first.

ACRL AND ARL STATISTICS

The recent publication of *ACRL University Library Statistics* for 1978–79, together with the annual issue of *ARL Statistics*, offers for the first time a body of timely and more or less comparable data on university libraries.[3]

Kendon Stubbs is associate university librarian, University of Virginia, Charlottesville.

These two compilations provide data on the libraries of 177 of the 181 U.S. institutions classified by the Carnegie Council as doctorate-granting institutions, as well as data on 19 Canadian university libraries—a total of 196 libraries.[4] There are twenty-two categories of library data concerning collections, interlibrary loans, expenditures, and personnel; data are also reported on enrollments, Ph.D.s awarded, and Ph.D. fields.

As we have seen, the *Standards for University Libraries* do not present quantitative criteria or levels of excellence to be used as measures of achievement. The ACRL and ARL data, therefore, will not reveal whether this or that library meets accepted quantitative standards. But in certain ways the data can tell us where university libraries are, if not where they should be. We cannot say that a university library has satisfied or failed to satisfy external criteria, but empirically we can describe the quantitative relationships among university libraries in 1978–79. This paper, therefore, discusses some ways in which the data can answer two kinds of questions:

1. What are the relationships among various categories of library and university data—for example, between the numbers of staff and the sizes of libraries in volumes held?

2. Is it possible to distinguish among groupings of libraries and to describe various groupings quantitatively?

Before turning to these questions, however, we need to be aware of two caveats about the ACRL and ARL data.

First, except for the categories concerning interlibrary loans, the data do not necessarily tell us anything about quality of service. It is true that if a scholar wants Kalkar's *Ordbog til det aeldre danske sprog*, no doubt a university library with more than a million volumes is the best place to try. On the other hand, it may be that one will find *Lolita* more easily in a community college than a university library. DeGennaro has pointed out that our statistics are merely measuring degrees of bigness, not availability or accessibility of information.[5] Some attempts are being made to relate size to service,[6] but we are not yet able to claim even feebly that the ACRL and ARL data disclose much about how well our users are served. In the terms of Brown's recent typology of information about libraries, the data provide measures of resources but not measures of library activities, users, or performance.[7]

Second, although the ACRL and ARL publications are undoubtedly the most useful statistical compilations on university libraries, remember that they are subject to the vagaries that willy-nilly beset data collections. Piternick claimed that the user of those data will rely upon them as the drunk relies upon the street lamp—for support rather than illumination.[8] Even if one does not take so tight a stand, it is at least worthwhile to follow Piternick's advice that the data ought to be handled with care. One need only glance through the seventeen pages of notes on the ten pages of data in the latest *ARL Statistics* to realize the variety in the bases for reporting data.[9] In short, the ACRL-ARL data can disclose much about library size and resources deployed, but not everything.

With these cautions in mind, we turn next to a discussion of the relationships among categories of data in university libraries.

Similarities in University Libraries: Correlation and Regression

Although the *Standards for University Libraries* avoid quantitative criteria, an appendix discusses "Quantitative Analytical Techniques for University Libraries." Among the techniques suggested are ratio analysis and regression analysis. We consider first the use of ratio analysis with the ACRL-ARL data.[10]

A table of various ratios is presented in both the *ACRL Statistics* (p.12) and the *ARL Statistics* (p.14). What is interesting about these two sets of ratios is how closely, for the most part, they correspond. In both the ACRL and ARL libraries, the median number of professional staff is 25 percent of total staff. In both, the median ratio of professionals to nonprofessionals is 0.5:1. Serials expenditures are 49 percent of library materials expenditures in the ACRL universities and 54 percent in the ARL. In the ACRL, 36 percent of total expenditures is for library materials, in the ARL, 31 percent. Only in the ratio of items loaned to items borrowed is there a striking difference: 1.5:1 in the ACRL, 2.4:1 in the ARL. It is tempting to assume that these ratios offer a firm ground for statements about university libraries—to conclude, for example, that in universities the ratio of nonprofessionals to professionals is two to one, that about one-third of library expenditures is for materials, that about 50 percent of the money for materials is committed to subscriptions and standing orders, and so on.

But even when we know the high, low, and median ratios, we have no measure of how closely the ratios for individual libraries cluster about the median. For instance, if a library spent 60 percent of its materials budget on serials, is that library significantly out of line with its peers? A measure of relative variability called the coefficient of variation can indicate the utility of the different ratios. This shows whether the values of a ratio in the individual libraries are fairly similar or more widely dispersed. As examples, in the ACRL-ARL data the median ratio of total salaries to total expenditures is .55:1, and the median ratio of nonprofessionals to professionals is 2:1. But the coefficient of variation for salaries to total expenditures is 13 percent, and for nonprofessionals to professionals 34 percent. The former ratio is considerably more informative than the latter. We come closer to conveying a quantitative truth about university libraries when we say that total salaries are about 55 percent of total expenditures than when we say that university libraries have two nonprofessionals for each professional. Ratio analysis is thus a useful starting point in analyzing data. But by itself it leaves us in the dark when we try to assert that this or that ratio is characteristic of university libraries. For a data analysis technique that indicates how accurate our assertions are likely to be, we must turn to correlation and regression analysis.

The appendix to the *University Standards* contains some comments on regression analysis, and there are descriptions in most statistics textbooks.[11] For the purposes of the following discussion it is worth noting that some of the basic concepts of regression can be grasped through reference to simple geometry. Suppose that we have two variables, or categories of data, such as volumes held and professional staff. If we plot the two variables on a graph (number of volumes along the x axis and professional staff along the y axis), each point will represent the professionals and volumes of one library. The straight line that lies closest to all of the points is the regression line. The general formula for a straight line in geometry is $Y = a + bX$. In our example, professionals (Y) $= a + b$ times volumes (X). Regression analysis calculates the values of a and b in the formula. Thus, in the most accurate way, the formula describes the linear relationship between two variables. How strong the relationship is (how close the points are to the regression line) is indicated by the coefficient of determination, r^2. If the points do not have any measurable relationship to the straight line, r^2 equals zero. If the points all lie exactly on the line, r^2 equals one. In different terms, r^2 measures how much of the variation in one variable is associated with the variation in the other. Where r^2 equals one, all of the variation in the first variable can be explained by reference to the second variable.

Consider again professionals and volumes held. For the 196 ACRL and ARL university libraries the regression equation is $Y = 11.84 + .0000274X$; or, prof. staff $= 11.84 + .0000274 \times$ volumes. Here r^2 equals .86. With a high degree of accuracy, the regression equation describes the relationship between volumes held and professional staff in university libraries. Eighty-six percent of the variation in the numbers of professionals can be accounted for by the volume sizes of the libraries. If we substitute 36,500 for X in the regression equation, $Y = 11.84 + .0000274 \times 36{,}500 = 11.84 + 1$. Consequently, for each 36,500 volumes, the equation predicts 11.84 (or approximately 12) plus one professional. If a library has 2,190,000 volumes, or 36,500 times 60, then the formula predicts that that library has 12 plus 60, or 72, professionals. The formula is a powerful tool for making a statement about a quantitative relationship in university libraries. It tells us that, in general, university libraries in 1978–79 had one professional for each 36,500 volumes held, added to a base of 12 professionals.

Note that the values predicted by the formula will rarely coincide precisely with the actual numbers of professionals, since the relationship between professionals and volumes is not perfect but rather is characterized by the r^2 of 86 percent. Some of the actual numbers of professionals will be less than the formula predictions and some greater. The difference between an actual and a predicted number of professionals is called a residual. Regression analysis offers a way of characterizing the relative size of individual residuals. For the regression of professionals with volumes, one standard deviation of the residuals is approximately 14. In general, we can expect that about two-thirds of the residuals will be between −14 and +14; and 95 percent of the residuals will fall between −28 and +28 (that is, two standard deviations).

In illustration of the foregoing discussion, consider two university libraries picked at random. In library *A*, volumes held are 513,036, and the actual number of professionals is 30. In library *B* volumes are 1,921,278 and professionals 43. Substituting the volume figures in the formula produces a prediction of 26 professionals for library *A* and 65 for *B*. The formula underpredicts *A* by 4 professionals and overpredicts *B* by 22. The residual 4 is well within one standard deviation of 14. Library *A* therefore exhibits a professional staffing fairly typical of university libraries. For library *B*, on the other hand, the residual of 22 is between one and two standard deviations, or between 14 and 28. In this case there is a question whether *B* is understaffed in relation to what is typical of professional staffing in university libraries. (It should be noted, however, that there may be local conditions that make the staffing of *B* right for its situation. The regression equation tells us that, when we consider size in volumes alone, most university libraries have actual professional staffs within about 14 above or below what the equation predicts. But the regression analysis does not consider the multitude of local influences on staff size.)

Just as we can show a relationship between volumes and professionals (one professional for each 36,500 volumes, above a base of 12), so we can discern other relations in the ACRL-ARL data. Some of these are displayed in table 1. The first entry in the table, for example, indicates that, over and above 13,600 gross volumes added, university libraries added one volume for every 33 volumes held. This formula has an associated r^2 of 78 percent. The standard deviation of the residuals (the differences between actual volumes added and added volumes predicted by the formula) is 20,800. (In table 1 the numbers in the regression equations and the standard deviations are rounded, for simplicity. "Total staff" equals professional plus nonprofessional staff.)

The r^2s in table 1 are the highest that can be achieved (and indeed are very respectable) when we use only one variable to predict another, unless we use less meaningful predictors. For instance, volumes added net will predict volumes added gross with an r^2 of 95 percent. But we do not come away much wiser from learning that, if we have such and such a number of net volumes added, we should have some number of gross volumes added. Is it possible otherwise to obtain higher r^2s than those in table 1? There are two ways to make the predictions more accurate. First, instead of using just one predictor, we can use two or more in the regression equation. As an example, we have used volumes held to predict professional staff, with an r^2 of 86 percent. Through multiple regression analysis we can predict professionals with the following variables in the equation: volumes held, volumes added gross, microforms, current serials, interlibrary loans and borrowing, total students, graduate students, Ph.D.s awarded, and Ph.D. fields. But here the R^2 is 90 percent—not significantly better than the 86 percent with volumes alone. It has been noted in the past that library variables are highly correlated with each other. The more volumes a library has, the more it has of serials, professionals, expenditures, and so on. As a result, it is hard to make a much better prediction of a variable like professionals with multiple predictors than we can get from one predictor like volumes, because the other predictors cannot add much to what volumes have already contributed.

A second possible method of improving the r^2s is to divide the ACRL-ARL libraries into smaller groups. This method is suggested by the appendix to the *University Standards*, following the procedure of Baumol and Marcus. We might, for example, consider the ACRL libraries separately from the ARL libraries. Or we might further divide these groups into public, private, and Canadian libraries, and subject each group to regression analysis. Space does not permit a display of the results of regression with these various groupings. Suffice it to say that, when regression analysis is carried out on these groups, in most cases

TABLE 1

REGRESSION RESULTS FOR SELECTED VARIABLES IN ACRL-ARL DATA, 1978–79

Variable Predicted	Regression Equation	r^2	Standard Deviation of Residuals
Volumes added, gross	1 for each 33 vols. held + 13,600	78%	20,800
Current serials	1 for each 92 vols. held + 1,000	84%	6,200
Expenditures for library materials	\$15 for each vol. added gross + \$360,000	77%	\$365,000
Total library expenditures	\$68,000 for each professional + \$290,000	91%	\$806,000
Professionals	1 for each 36,500 vols. held + 12	86%	14
Total staff	1 for each 11,800 vols. held + 37	81%	54

the r^2s do not differ significantly from the r^2s of the entire ACRL-ARL. The only groups that do display significant differences are the ACRL, where the r^2s are lower, and the private universities, where the r^2s are higher. (These results point to more variability in the ACRL libraries than in the whole group of universities, whereas the private institutions show greater homogeneity.)

Can the regression equations of table 1, or other regression results, be taken as quantitative standards for university libraries? Can we say that above certain bases university libraries ought to add one volume for each 33 volumes held and spend \$15 per volume, that they should have one staff member for each 11,800 volumes held, and that total expenditures should amount to \$68,000 for each professional on the staff? Not really. These equations merely indicate what was characteristic of university libraries in 1978–79. They do not tell us whether the resources of the libraries were able to provide as well as possible for the needs of their users. The equations do not permit us to make the leap from what is to what should be.

The equations, moreover, do not necessarily characterize university libraries as distinct from other kinds of libraries. Consider again the equation linking professionals with volumes: one professional for each 36,500 volumes held, plus 12 professionals. When regression analysis is performed on the 1976–77 HEGIS data for the approximately 3,000 academic libraries in the United States, it turns out that the equation for all academic libraries is: one professional for each 34,800 volumes held, plus two professionals, with an r^2 of 85 percent. Except for the base of 12 or 2 professionals, there is little difference between the equations for the university libraries and for the entire population of 3,000 U.S. academic libraries. Above a certain base, all college and university libraries seem to have had approximately one professional for each 35,000 volumes. The regression equations of table 1 consequently cannot serve as standards peculiar to university libraries.

In the remainder of this paper we shall consider some of the methods by which university libraries can be differentiated from other libraries, and by which various levels of university libraries can be distinguished.

GROUPINGS OF UNIVERSITY LIBRARIES: DISCRIMINANT ANALYSIS

If we look through the ACRL and ARL data, it is hard to find gaps in the range of data from the smallest library to the largest. Most observers would probably decide that Harvard, at one end of the scales, and possibly U.S. International, the New School, and Rockefeller, at the other end, are somehow different from the other libraries. But between these extremes one finds no quantum jumps from one state of university library to another. Yet it is possible quantitatively to distinguish one kind of university library from another—to find, in other words, that there are statistically distinct groupings among the libraries.

In the investigation of groupings a useful tool is the statistical technique called discriminant analysis.[12] Discriminant analysis begins with two or more discrete groups—for instance, male and female library professionals. It then analyzes discriminating variables—e.g., salaries, salary increases, rank—to determine which combinations of the data best distinguish between the groups. A result of the analysis is a formula by which, in the present example, we can differentiate males from females on the basis of their salaries, raises, and rank. Once we have the formula, we can use it to classify individuals as male or female. We can then see how much discriminating power the formula offers. It is interesting to note that in university libraries a discriminant formula can sometimes correctly classify 75–85 percent of professionals as males or females merely by reference to their salaries and raises—an indication of the salary differentials between men and women in libraries.

For present purposes, perhaps the first obvious question to put to discriminant analysis is whether the ACRL libraries comprise a group statistically distinct from the ARL libraries. We need to test a set of variables to determine whether some combination of the variables can discriminate between the ACRL and ARL. Previous analysis has shown that, of the twenty-two variables reported in the *ARL Statistics*, only ten are necessary to characterize library size and resources deployed.[13] This analysis has been replicated for the ACRL-ARL data with the same result. The ten variables are:

volumes held
volumes added, gross
microforms
current serials
expenditures for library materials
expenditures for binding
total salaries
other operating expenditures
professional staff
nonprofessional staff

These ten variables can therefore be used as the discriminating variables.[14]

Discriminant analysis finds that the greatest differentiation between the ACRL and ARL occurs when five variables are in the discriminant equation: volumes held, volumes added gross, microforms, expenditures for library materials, and professional staff. The equation based on these five variables correctly classifies 94 percent of the libraries as either ACRL or ARL. Only five ARL libraries are misclassified as ACRL, and six ACRL libraries as ARL. Discriminant analysis thus tells us that there is a remarkably strong statistical distinction between the ACRL and ARL libraries. If we have a few items of data from a university library—volumes held, volumes added, microforms, and so on—we can predict with 94 percent certainty whether that library belongs to the ARL or ACRL.

Are there any other discrete groups that allow similar accuracf of classification? Another obvious set to try is the Carnegie Classification groups. The ACRL-ARL data are for the libraries of those institutions termed doctorate-granting institutions by the Carnegie Council. The council further subdivides these institutions into research universities and doctorate-granting universities. Can we use library data to distinguish between these two kinds of universities? The answer from discriminant analysis is that only 80 percent can be classified correctly. That is, from library data we can predict with only 80 percent certainty whether parent institutions are research or doctorate-granting universities. Similarly, library data permit us to classify correctly as public or private only 75 percent of the institutions. Other possible groupings are based on enrollments or degrees awarded or Ph.D. fields. Can library variables distinguish between institutions with greater and lesser numbers of graduate students? In other words, is there a correspondence between library size and number of graduate students? We can divide the 196 ACRL-ARL institutions into two groups with median enrollments, Ph.D.s awarded, or Ph.D. fields as the dividing points between the groups. Then we can use the library variables to determine how distinct the groups are. The results from discriminant analysis are all significantly lower than the 94 percent correct classification of libraries as ACRL or ARL.*

*Segmenting the data of a continuous variable like enrollments and then performing discriminant analysis on the resulting groups is a procedure open to some criticism. It is followed here merely because ɛt points simply to some basic results that can be confirmed by more abstruse statistical techniques.

These results are not surprising. Over the years the chief criterion for ARL membership has been library size, and so the distinction between the ARL and ACRL is based on library variables. The distinction between other groups like the Carnegie groups is based on university variables. Library variables are much more closely correlated with one another than with measures of university size, like enrollments and degrees awarded. Through a statistical technique known as canonical correlation we can compare the ten library size variables with the university size variables. It turns out that at most 78 percent of the variation in library size is associated with variation in university size, and vice versa. Up to a point we can understand library size by examining the parent institutions, but about one-quarter of the variation in library size cannot be accounted for by university data. We find, moreover, that the strongest relations are between library size and graduate enrollments, and to a lesser extent, Ph.D.s awarded. Total students and Ph.D. fields have little relation to library size. The college library standards relate collection size and library personnel to numbers of students and faculty. For university libraries, however, there are statistical reasons why library variables concerning collections, expenditures, and staff need to be related to each other, rather than to university data.

Discriminant analysis thus points to the following conclusions. There is a strong statistical distinction between ARL and ACRL libraries. This distinction is firmer than that between other groups based on university characteristics such as enrollments or degrees awarded. From library data we can tell whether a given library is part of the ARL or ACRL, but we cannot tell as much about the university to which the library belongs. Should we conclude further that the ARL group represents a different kind of library from the ACRL? The answer must be no. As shown at the beginning of this section, in the entire range of ACRL-ARL data there are no obvious jumps from one level to another. The ACRL merges into the ARL. Discriminant analysis allows us to say that ARL libraries as a whole are distinct from ACRL. What is needed is a method of determining how similar individual ACRL libraries are to ARL and vice versa. The following section examines this problem.

Differences among University Libraries: Principal Component Analysis

The preceding analysis suggests that it is valid to measure the quantitative characteristics of either the ACRL or the ARL libraries and then to compare the libraries of the other group by these measurements. The technique that we shall use for these comparisons is principal component analysis, a variant of the statistical procedures called factor analysis.[15]

Principal component analysis begins with a set of variables such as the ten library size variables listed above. It derives a weight, or component score coefficient, for each variable according to how similar or dissimilar the libraries are in respect to that variable. For example, the ACRL-ARL libraries are most alike in the total salaries they pay, and consequently total salaries have the highest component score coefficient. The libraries exhibit the greatest variability in microforms, which have the lowest weight. These coefficients or weights are then multiplied by the data for each library to produce a component score for that library. The scores thus represent no more than a sum of the data from each library on its collections, expenditures, and staffing, weighted in accord with the ways in which the libraries are similar or different. They are simply mathematical transformations of the data for each library.

It is interesting, however, that as a whole the scores are approximated by a standard normal curve or a bell-shaped curve. In this kind of curve or distribution the midpoint (that is, the mean and the median) is zero. Most of the values fall between $+2$ and -2, a distribution that permits useful probability statements. For example, in any standard normal distribution approximately 84 percent of the values is greater than -1, 95 percent than -1.65, and 99 percent than -2.33.

We can use the probability feature of the component scores to describe similarities and differences among the ACRL and ARL. Suppose that we calculate scores for the ARL. Then the whole range of scores indicates ARL library size and resources deployed. If a library shares the essential quantitative features of the ARL members, the chances are 95 percent that the component score for that library will be above −1.65, and 99 percent that it will be above −2.33. In different terms, there is only a 1 percent probability that a library similar to the ARL libraries will score below −2.33.

In illustration, we compute component scores for the ARL and then, using the same formula, calculate scores for the ACRL libraries. These scores are displayed in table 2. Note that the scores for ARL libraries range from 3.05 to −1.91, in an approximately normal distribution, and the ACRL scores from −.42 to −7.17. Forty-seven ACRL libraries score lower than −2.33.

How should these scores be interpreted? In

TABLE 2
PRINCIPAL COMPONENT SCORES OF UNIVERSITY LIBRARIES, 1978–79 (FROM ARL COMPONENT SCORE FORMULA)

Library	Group	Score	Library	Group	Score
1. Harvard	ARL	3.05	50. South Carolina	ARL	-.32
2. Calif., Berkeley	ARL	2.18	51. Connecticut	ARL	-.33
3. Yale	ARL	2.12	52. Syracuse	ARL	-.34
4. Indiana	ARL	1.97	53. Missouri	ARL	-.35
5. Calif., Los Angeles	ARL	1.92	54. Johns Hopkins	ARL	-.35
6. Toronto	ARL	1.91	55. Tennessee	ARL	-.36
7. Illinois	ARL	1.88	56. M.I.T.	ARL	-.39
8. Stanford	ARL	1.80	57. Western Ontario	ARL	-.39
9. Washington	ARL	1.70	58. Washington U–St. Louis	ARL	-.40
10. Texas	ARL	1.62	59. Utah	ARL	-.40
11. Michigan	ARL	1.62	60. Wayne State	ARL	-.41
12. Columbia	ARL	1.54	61. Laval	ACRL	-.42
13. Cornell	ARL	1.47	62. Nebraska	ARL	-.51
14. Wisconsin	ARL	1.40	63. Arizona State	ARL	-.51
15. Minnesota	ARL	1.03	64. Temple	ARL	-.52
16. British Columbia	ARL	.96	65. Louisiana State	ARL	-.52
17. Chicago	ARL	.90	66. Texas A&M	ARL	-.53
18. North Carolina	ARL	.87	67. York	ARL	-.56
19. Rutgers	ARL	.83	68. Purdue	ARL	-.56
20. Florida	ARL	.76	69. Cincinnati	ARL	-.56
21. Virginia	ARL	.72	70. Iowa State	ARL	-.56
22. Princeton	ARL	.72	71. Boston	ARL	-.58
23. Pennsylvnia State	ARL	.66	72. Joint University	ARL	-.60
24. Northwestern	ARL	.63	73. Brigham Young	ARL	-.65
25. Ohio State	ARL	.59	74. SUNY–Stony Brook	ARL	-.67
26. Pennsylvania	ARL	.54	75. Emory	ARL	-.67
27. Calif., Davis	ARL	.51	76. Ottawa	ACRL	-.71
28. New York	ARL	.46	77. Colorado	ARL	-.71
29. Alberta	ARL	.40	78. Massachusetts	ARL	-.72
30. Southern California	ARL	.30	79. Rochester	ARL	-.72
31. Pittsburgh	ARL	.29	80. Georgetown	ARL	-.72
32. Georgia	ARL	.29	81. Miami	ARL	-.73
33. Michigan State	ARL	.27	82. Calif., Irvine	ACRL	-.81
34. Duke	ARL	.26	83. Calgary	ACRL	-.81
35. SUNY–Buffalo	ARL	.21	84. Howard	ARL	-.82
36. Iowa	ARL	.19	85. Manitoba	ACRL	-.86
37. Arizona	ARL	.17	86. Brown	ARL	-.89
38. Houston	ARL	.14	87. Oklahoma	ARL	-.90
39. Kansas	ARL	.11	88. Queens	ARL	-.91
40. Maryland	ARL	.08	89. Oregon	ARL	-.91
41. McGill	ARL	.03	90. North Carolina State	ACRL	-.95
42. Calif., San Diego	ARL	.02	91. New Mexico	ARL	-.97
43. Southern Illinois	ARL	-.03	92. Waterloo	ACRL	-.97
44. Kentucky	ARL	-.03	93. Calif., Riverside	ARL	-.99
45. Hawaii	ARL	-.11	94. Carleton	ACRL	-1.05
46. VPI & SU	ARL	-.12	95. SUNY–Albany	ARL	-1.05
47. Calif., Santa Barbara	ARL	-.17	96. McMaster	ARL	-1.05
48. Florida State	ARL	-.20	97. Wisconsin, Milwaukee	ACRL	-1.07
49. Washington State	ARL	-.31	98. Dartmouth	ARL	-1.13

TABLE 2 (CONTINUED)

Library	Group	Score	Library	Group	Score
99. Colorado State	ARL	-1.14	148. Rhode Island	ACRL	-2.54
100. Tulane	ARL	-1.21	149. Utah State	ACRL	-2.54
101. Case Western Reserve	ARL	-1.22	150. Northeastern	ACRL	-2.55
102. Guelph	ARL	-1.24	151. St. John's	ACRL	-2.57
103. Auburn	ACRL	-1.26	152. Tufts	ACRL	-2.67
104. Notre Dame	ARL	-1.28	153. Wyoming	ACRL	-2.68
105. Northern Illinois	ACRL	-1.33	154. Catholic	ACRL	-2.68
106. Alabama	ARL	-1.39	155. Brandeis	ACRL	-2.69
107. Illinois, Chicago Circle	ACRL	-1.53	156. Tulsa	ACRL	-2.72
108. West Virginia	ACRL	-1.57	157. Texas Christian	ACRL	-2.82
109. Delaware	ACRL	-1.59	158. Adelphi	ACRL	-2.89
110. Kent State	ARL	-1.60	159. Northern Colorado	ACRL	-2.91
111. Ball State	ACRL	-1.62	160. Alaska, Fairbanks	ACRL	-2.98
112. Georgia Inst. of Tech.	ACRL	-1.62	161. Lehigh	ACRL	-3.05
113. Oregon State	ACRL	-1.70	162. Idaho	ACRL	-3.08
114. Illinois State	ACRL	-1.73	163. East Texas State	ACRL	-3.10
115. Fordham	ACRL	-1.74	164. William and Mary	ACRL	-3.13
116. Virginia Commonwealth	ACRL	-1.75	165. Maine, Orono	ACRL	-3.21
117. South Florida	ACRL	-1.76	166. Southern Mississippi	ACRL	-3.28
118. Louisville	ACRL	-1.79	167. South Dakota	ACRL	-3.40
119. Georgia State	ACRL	-1.82	168. Montana	ACRL	-3.54
120. Texas Tech.	ACRL	-1.84	169. American	ACRL	-3.63
121. Oklahoma State	ARL	-1.87	170. Montana State	ACRL	-3.66
122. Rice	ARL	-1.91	171. North Dakota	ACRL	-3.69
123. Simon Fraser	ACRL	-1.93	172. Texas Woman's	ACRL	-3.75
124. North Texas State	ACRL	-1.94	173. Calif. Inst. Of Tech.	ACRL	-3.80
125. Miami (Ohio)	ACRL	-1.96	174. Detroit	ACRL	-3.82
126. Southern Methodist	ACRL	-2.04	175. Idaho State	ACRL	-4.04
127. Nevada, Reno	ACRL	-2.05	176. Rensselaer Polytechnic	ACRL	-4.28
128. Memphis State	ACRL	-2.08	177. Carnegie-Mellon	ACRL	-4.56
129. Akron	ACRL	-2.09	178. South Dakota State	ACRL	-4.58
130. Calif., Santa Cruz	ACRL	-2.11	179. Clark	ACRL	-4.80
131. New Hampshire	ACRL	-2.20	180. Pacific	ACRL	-5.10
132. Claremont	ACRL	-2.20	181. Missouri, Rolla	ACRL	-5.40
133. Vermont	ACRL	-2.23	182. Illinois Inst. Of Tech.	ACRL	-5.75
134. Arkansas, Fayetteville	ACRL	-2.27	183. New School	ACRL	-6.44
135. Toledo	ACRL	-2.28	184. Rockefeller	ACRL	-6.50
136. New Mexico State	ACRL	-2.31	185. United States Intl	ACRL	-7.17
137. Denver	ACRL	-2.32	186. Kansas State	ACRL	*
138. Mississippi State	ACRL	-2.33	187. Mississippi	ACRL	*
139. Bowling Green State	ACRL	-2.36	188. Montreal	ACRL	*
140. Clemson	ACRL	-2.37	189. New Bruns., Fredericton	ACRL	*
141. George Washington	ACRL	-2.41	190. North Dakota State	ACRL	*
142. Indiana State	ACRL	-2.45	191. Ohio	ACRL	*
143. North Carolina, Grnsboro	ACRL	-2.45	192. St. Louis	ACRL	*
144. Missouri, Kansas City	ACRL	-2.46	193. SUNY–Binghamton	ACRL	*
145. Loyola, Chicago	ACRL	-2.52	194. Western Michigan	ACRL	*
146. Marquette	ACRL	-2.52	195. Windsor	ACRL	*
147. Hofstra	ACRL	-2.53	196. Yeshiva	ACRL	*

*Missing data for these libraries preclude the calculation of component scores.

statistics it is customary to take a 95 or 99 percent cutoff point for rejecting a given hypothesis. In the present case we might select the more inclusive 99 percent, with a corresponding score of -2.33. Then we should say that libraries that score below -2.33 probably do not share the library size characteristics of the ARL libraries. Statistically, it is likely that the libraries with scores above -2.33 are a different kind of library from those with scores below -2.33. What is characteristic of the ARL libraries in collections, staffing, and expenditures is shared by 138 university libraries with scores above -2.33, but is lacking in the 47 libraries with scores below -2.33. This number, -2.33, therefore serves as a minimum threshold for the majority of university libraries.

The component scores are a sum of the data for ten variables. Consequently, different combinations of data can produce the same score. One library that has, for example, a large number of volumes and few serials can have the same score as another library with fewer volumes but more current serials. To provide a clearer picture of what the −2.33 threshold implies, however, we can mathematically transform −2.33 into a value for each of the ten variables. These transformations are shown in table 3.

The "dividing lines" of table 3 can be interpreted in this way: If the numbers of volumes held in ARL libraries are transformed into approximately a standard normal distribution, a value of −2.33 corresponds to 600,000 volumes. We should expect that 99 percent of libraries like the ARL libraries would have 600,000 volumes or more. When we find 39 libraries (20 percent of all university libraries) with fewer than 600,000 volumes, we have to conclude that these are statistically different in kind from the ARL-like university libraries in respect to numbers of volumes held. Thus, 600,000 volumes serves as a minimum, dividing the major group of university libraries from the other libraries; and similarly for the other nine variables. It would be wrong to argue that the 39 libraries with fewer than 600,000 volumes are somehow not university libraries. They are, in fact, as much as the other 157, the libraries of institutions classified by the Carnegie Council as universities. What can be concluded, however, is that from a statistical standpoint there is an overriding probability that a library must have at least 600,000 volumes in order to share the essential quantitative characteristics of most university libraries.

In arriving at these conclusions, we began by using the ARL libraries as a base from which to measure university library characteristics. Obviously, we could in the same ways use the ACRL as a base. In this case the rank order of libraries in table 2 would remain about the same. But approximately the first 34 libraries (from Harvard through Duke) would have scores greater than 2.33. We should then say that these 34 libraries are statistically different from the other ARL and ACRL libraries. But it is not clear what this statement would imply: that there are university libraries, and then there are some 30 superlibraries? The implications of table 2 seem more reasonable: that most university libraries, from Harvard through ACRL libraries, share the same kinds of quantitative characteristics; but libraries in the lower end of this range increasingly assume the features of smaller institutions, such as college libraries.

University Library Standards?

Tables 1 and 3 together offer what seems very much like quantitative standards for university libraries. For example, table 1 shows that the typical university library has

TABLE 3

99 Percent (−2.33) Approximate Dividing Lines for University Library Variables, 1978–79

		No. of Libraries below Dividing Line			
		ACRL		ARL	
Variable	Dividing Line	No.	%	No.	%
Volumes	600,000	39	40%	0	0%
Volumes added, gross	24,000	38	41%	1	1%
Microforms	425,000	54	55%	3	3%
Current serials	6,000	36	38%	0	0%
Expenditures for library materials	$620,000	41	42%	0	0%
Expenditures for binding	$ 30,000	32	33%	1	1%
Total salaries	$890,000	37	38%	0	0%
Other operating expenditures	$110,000	43	45%	1	1%
Professionals	23	36	37%	0	0%
Nonprofessionals	46	47	48%	0	0%

TABLE 4

MINIMAL LEVELS FOR UNIVERSITY LIBRARIES, 1978–79

Category	At Least Equal To:	And No Fewer Than:
Volumes held	— —	600,000
Volumes added, gross	Vols./33 – 7,200	24,000
Current serials	Vols./92 – 5,200	6,000
Expenditures for library materials	Vols. added gross x $15 – $5,000	$620,000
Total library expenditures	Profs. x $68,000 – $516,000	$1,650,000
Professionals	Vols./36,500 – 2	23
Total staff	Vols./11,800 – 17	69

twelve professionals plus one professional for each 36,500 volumes held. In most ACRL-ARL libraries the actual staffing is within fourteen professionals of what this formula predicts. The formula prediction minus fourteen is therefore a minimum for most university libraries. That is, professionals equal volumes divided by 36,500, plus twelve, minus fourteen, or vols./36,500 – 2. From table 3 the typical university library has at least twenty-three professionals. We can therefore say that, as a minimum, the number of professionals needs to be (1) at least equal to volumes/36,500 – 2 and (2) no less than twenty-three. Table 4 displays some of these minima.* On the average about 10 percent of the ARL libraries and 38 percent of the ACRL, or 25 percent of all university libraries, are below each of these levels.

Are the minimal levels of table 4 at last the elusive quantitative standards for university libraries? Certainly they are empirical criteria that point to what was characteristic of university libraries in 1978–79. We might even say that, if a library does not want to fall below 1978–79 university library levels, it must satisfy the criteria of table 4. But the criteria in a way represent the lowest permissible statistical thresholds. The 75 percent of university libraries that have surpassed these lower limits would rightly feel cheated (or worse) if they were told that they could have expenditures for library materials equal to only $15 per volume added, minus $5,000, or professionals equal to only vols./36,500 – 2. These are not standards in the sense of goals that most libraries should strive to achieve. More importantly, the criteria also fail to reveal whether the collections, expenditures, and staffing of table 4 are sufficient "to support the university's total instructional needs and to facilitate the university's research programs."[16] We have not yet arrived at a means of comparing these criteria with measures of library activities, users, and performance.

At this point one may feel somewhat like the dreamer of *Piers Plowman*, who through 7,303 lines of poetry seeks for what he should do to win salvation, and in the end learns that the search must begin again. University libraries that wonder what they ought to do to be saved will not find the answers in table 4. They must look for and measure what is necessary to give users what they need when they need it. But that search will be considerably more arduous and time-consuming than the one described here.

*For the figures from table 3 total library expenditures equal expenditures for library materials plus binding plus total salaries plus other operating expenditures. Total staff equals professionals plus nonprofessionals.

REFERENCES

1. "Standards for University Libraries," *College & Research Libraries News* 40:101–10 (April 1979).
2. "Standards for College Libraries," *College & Research Libraries News* 36:277–79, 290–95, 298–301 (Oct. 1975).
3. *ACRL University Library Statistics 1978–79* (Chicago: Association of College and Research Libraries, 1980); *ARL Statistics 1978–79* (Washington, D.C.: Association of Research Libraries, 1979). The academic library data collected through the federal HEGIS surveys

are not issued in as timely a manner: at the time of writing, the latest data available even on computer tape from the National Center for Education Statistics were for 1976–77 (although the 1978–79 data became available in January 1981).

4. Carnegie Council on Policy Studies in Higher Education, *A Classification of Institutions of Higher Education* (rev. ed.; Berkeley: The Council, 1976). The classification actually counts 184 doctorate-granting institutions, because Boston, Cornell, and Joint University each have constituent parts listed in two categories.
5. Richard DeGennaro, "Library Statistics & User Satisfaction: No Significant Correlation," *Journal of Academic Librarianship* 6:95 (May 1980). See also Stella Bentley, "Academic Library Statistics: A Search for a Meaningful Evaluative Tool," *Library Research* 1:143–52 (Summer 1979).
6. For example, Paul B. Kantor, *Levels of Output Related to Cost of Operation of Scientific and Technical Libraries: The Final Report of the LORCOST Libraries Project* (Cleveland: Case Institute of Technology, 1980).
7. Maryann K. Brown, "Library Data, Statistics, and Information," *Special Libraries* 71:475–84 (Nov. 1980).
8. George Piternick, "ARL Statistics—Handle with Care," *College & Research Libraries* 38:420 (Sept. 1977).
9. *ARL Statistics 1979–80* (Washington, D.C.: Association of Research Libraries, 1980), p.48–64.
10. "Standards for University Libraries," p.107–10. For the sake of comparability in the following analyses, the Canadian expenditures reported in the ACRL and ARL *Statistics* are converted to U.S. dollar equivalents at the rate of 1.1666 Canadian dollars to one U.S. dollar, from *Bank of Canada Review*, p.S113 (Oct. 1979).
11. A useful introduction to correlation and regression is Frederick Herzon, *Introduction to Statistics for the Social Sciences* (New York: Crowell, 1976), p.325–73. For an explanation in terms of university library data see Kendon Stubbs, *The ARL Library Index and Quantitative Relationships in the ARL* (Washington, D.C.: Association of Research Libraries, 1980), p.1–8. Examples of regression analysis of library data are William Baumol and Matityahu Marcus, *Economics of Academic Libraries* (Washington, D.C.: American Council on Education, 1973), p.20–40; and Donald Koepp et al., *Regression Analysis of the ARL Data* (Washington, D.C.: Association of Research Libraries, 1978).
12. Some introductions to discriminant analysis are Spencer Bennett and David Bowers, *An Introduction to Multivariate Techniques for Social and Behavioural Sciences* (New York: Wiley, 1976), p.95–117; David Kleinbaum and Lawrence Kupper, *Applied Regression Analysis and Other Multivariable Methods* (North Scituate, Mass.: Duxbury, 1978), p.414–46; and William Klecka, "Discriminant Analysis," in Norman Nie et al., *SPSS: Statistical Package for the Social Sciences* (2d ed.; New York: McGraw-Hill, 1975), p.434–67.
13. Stubbs, *ARL Library Index*, 9–10.
14. Because of the lognormal tendencies of the ACRL-ARL data, we use logarithms of the data, rather than the raw data, in these statistical analyses. For comments on the lognormal nature of library data see Allan Pratt, "The Analysis of Library Statistics," *Library Quarterly* 45:275–86 (July 1975).
15. Among the introductory treatments of principal component and factor analysis are Bennett and Bowers, *Introduction*, p. 8–71; R. J. Rummel, *Applied Factor Analysis* (Evanston, Ill.: Northwestern Univ. Pr., 1970); Jae-On Kim and Charles Mueller, *Introduction to Factor Analysis* (Beverly Hills, Calif.: Sage, 1978) and Jae-On Kim and Charles Mueller, *Factor Analysis: Statistical Methods and Practical Issues* (Beverly Hills, Calif.: Sage, 1978). The following discussion is based in part on Stubbs, *ARL Library Index*, p. 10–13. See also *ARL Statistics 1979–80*, p.23–25.
16. "Standards for University Libraries," p.102.