

WILLIAM E. McGRATH, RALPH C. HUNTSINGER, AND  
GARY R. BARBER

# An Allocation Formula Derived from a Factor Analysis of Academic Departments

*The authors derive a book fund distribution formula from a factor analysis of twenty-two variables which measure and quantify academic departments. The analysis generates a 22 x 22 matrix of correlations. A few of the significant correlations are discussed; e.g., those between books published and books circulated (high correlation) and circulation-by-subject and circulation-by-person (low correlation). The factor analysis sorts out the complex relationships between the twenty-two variables and reduces them to three main factors—two of which seem to describe materials used and users. The third may describe needs. The three factors are the chief elements in the formula. Each factor can be represented by any one or more of the variables in that factor.*

## PART I

### IDENTIFICATION OF VARIABLES AND COLLECTION OF DATA

COLLEGE AND UNIVERSITY libraries have many departments, institutes, and divisions competing for available library funds. Every librarian therefore has had to decide whether to: (1) emphasize and build one or more departments or divisions to the neglect of others; (2) assert no control and let a library collection develop where it may; or (3) emphasize all areas, fairly and equitably.

Too often the first two systems have

---

*Mr. McGrath is Head Librarian, University of Southwestern Louisiana, Dr. Huntsinger is Associate Professor of Chemical Engineering, and Mr. Barber is a graduate student in Chemical Engineering at South Dakota School of Mines and Technology, Rapid City.*

prevailed. The third has been tried, but few can agree on how to act "fairly and equitably." An objective, scientific technique for shaping the library's collection has never been developed. Ideally, a simple mathematical formula with as few variables as possible would be most desirable. The formula would be used to allocate the library's book budget to academic departments. Nearly every librarian allocates in one way or another. Even when he does not formally allocate with specific dollar amounts, he may subjectively allocate according to his own biases. If his bias is for chemical engineering, close study of the collection may reveal an unusually good chemical engineering section.

A good formula has been sought for

years. Formulas cited in the literature<sup>1</sup> are generally unsatisfactory. Most have been arbitrary, or based on what had been done in the past, or have not accounted for real and current needs. A few librarians, Ramer,<sup>2</sup> for example, have used many of the important elements in a formula but apparently without statistical justification. Richards<sup>3</sup> has mentioned the "continuing interest among the four out of five librarians who practice allocation today."

A good formula would help guarantee that available book funds will be distributed efficiently and equitably, that departments will be properly funded, and that the book collection will appropriately reflect the curriculum. In an effort to attain such a formula, the present study identified the forty-three variables which are defined and listed below in their naturally occurring groups. Some have been taken from Lyle<sup>4</sup> and other authors. Some are new. Some are simply derivatives of others—for example, G-1 is total inter-library loans while G-2 through G-9 are aspects of total inter-library loans. Each variable is a definition of "department."

McGrath<sup>5</sup> explains how a department can be defined as if it were a subject. Variables A-1 through A-3, B-1, F-8 and G-10 through G-12 define departments as subjects. All other variables define departments as organizations; *i.e.*, the number of people, credit hours, and so on.

<sup>1</sup> Guy R. Lyle, *Administration of the College Library* (3d ed. New York: H. W. Wilson, 1961), p. 348-49.

<sup>2</sup> James D. Ramer and Joseph Boykin, "The Book Budget in Academic Libraries," *Southeastern Librarian*, XVI (Spring 1966), 40-43.

<sup>3</sup> James H. Richards, Jr., "Academic Budgets and Their Administration," *Library Trends*, XI (April 1963), 415-26.

<sup>4</sup> Lyle, *op. cit.*

<sup>5</sup> William E. McGrath, "Determining and Allocating Book Funds for Current Domestic Buying," *CRL*, XXVIII (July 1967), 269-72.

TABLE 1

VARIABLES TO BE CONSIDERED IN A BOOK ALLOCATION FORMULA

A. *Books published.*

The total number of books published world-wide would be a desirable variable, but would be difficult to measure. For this project we tallied only those published in the U.S. The totals were derived from the several recent cumulations of the *American Book Publishing Record*.

1. Books published, total number.
2. Books published, total cost.
3. Books published, average cost.

B. *Existing collection.*

The existing collection can be counted item by item, but a linear measurement of the shelflist (100 books per inch of cards) is quicker.

The number of dollars allocated or spent, or the number of books bought in the immediate past should not be used in a formula because current conditions will be different. This is especially true if past buying and allocating was subjective. But it might be interesting to see how they correlate with other variables.

1. Relative strength of book collection.
2. Last year's departmental allocations, or expenditures.
3. Number of books purchased last year.

C. *Faculty and faculty load.*

The number of faculty in a department is a legitimate measure of its need. More difficult to justify, as a variable in a formula, is the length of time a person has been on the faculty. It is fair to assume that the longer a person has been around, the more likely it is that his *basic*

book needs have been filled, and that his years on the staff should rightly be scored *against* his department.

"Contact hours" are the number of class hours and laboratory hours a teacher actually spends with his students. "Equated hours" are a means of comparing contact hours to a norm, and are thus derivatives of contact hours. C-3, C-5, and C-7 are derivatives of C-4, C-6, and C-8, which in turn are total faculty hours per department. The assumption is that the greater the teaching load, the greater the book need.

"Faculty member" should include professors (full, associate, and assistant), instructors and, if desired, teaching assistants.

1. Number of faculty members in each department, instructors through full professors.
2. Faculty tenure (total number of years members have been on faculty).
3. Credit hours being taught—average per faculty member.
4. Credit hours being taught—totals. (Note under D-4, below.)
5. Contact hours—average per faculty member.
6. Contact hours—totals.
7. Equated hours—average per faculty member.
8. Equated hours—totals.

D. *Credit hours.*

One opinion is that no matter how many faculty members are in a department, or what their teaching loads are, what really counts is the number of courses offered and that the library is obliged to back up the courses with reading material whether or not the courses are actually taught in any given semester or year. Since a one-credit course

cannot be equated to a two, three, or four-credit course, the best way to consider them is according to the total number of credits per department.

Credits can be taken from the college catalog, and from changes on file in the Registrar's or Admissions Office. As shown, credits can be counted several ways. Credits for courses taught in two or more semesters per year can be counted more than once or only once, and credits "to be announced" (TBA's) counted as three each, or otherwise, as desired.

1. Credit hours, undergraduate, offered or listed.
2. Credit hours, graduate, offered or listed.
3. Credit hours, graduate and undergraduate, offered or listed.
4. Credit hours of courses actually being taught. (All sections counted. Counting TBA's as 3 each. This is the same as C-4, above, except that C-4 does not include TBA's.)
5. Credit hours of courses actually being taught. (Not more than one section counted for each course.)

E. *Enrollment.*

It is wise to use official figures whenever possible. Registrar's offices tally enrollment in several ways. Most use an official definition of "full-time equivalent" student. Enrollment here means "majors." Therefore, total enrollment which includes general students, specials, and non-declared majors cannot be used. More meaningful than "majors" perhaps is the number of students taking courses in "major" departments. This variable is actually tabulated in C-5 and C-6 above.

1. Enrollment, graduate and un-

dergraduate together.

2. Enrollment, graduate only.
3. Enrollment, undergraduate only.

#### F. Circulation.

Circulation, to be considered, must somehow be made relevant to departments. Two methods for doing so are (1) circulated books tabulated according to the borrower's departmental affiliation, as in F-5 through F-7, and (2) books circulated according to their relevance to a department's subject, as in F-8. Another paper by McGrath<sup>6</sup> on this subject explains how circulation according to department/subject can be tabulated.

Circulation from department libraries might be a problem and should be counted if possible.

1. Circulation, faculty and students—books and periodicals, plus inter-library loans. (F-2 plus G-1).
2. Circulation, faculty and students—books and periodicals. (F-3 plus F-5).
3. Circulation, graduate and undergraduate—books and periodicals. (F-4 plus F-6).
4. Circulation, undergraduate—books and periodicals.
5. Circulation, faculty—books and periodicals.
6. Circulation, graduate—books and periodicals.
7. Circulation, faculty and graduate—books only, plus inter-library loans (periodicals and books). (F-8 plus G-1).
8. Cumulative circulation count by department/subject.

#### G. Inter-library loans.

Inter-library loans, like circulation,

can also be counted according to department affiliation (G-1 through G-9) or according to subject (G-10 through G-12).

1. Inter-library loans, faculty and students—periodicals and books.
2. Inter-library loans, students—periodicals and books.
3. Inter-library loans, faculty—periodicals and books.
4. Inter-library loans, faculty and students—journals.
5. Inter-library loans, faculty and students—books.
6. Inter-library loans, students—journals.
7. Inter-library loans, students—books.
8. Inter-library loans, faculty—journals.
9. Inter-library loans, faculty—books.
10. Inter-library loans by subject, periodicals and books together.
11. Inter-library loans by subject, periodicals.
12. Inter-library loans by subject, books.

#### H. References in theses.

The assumption is that references, because they are *cited*, indicate their true value. Many books are circulated for graduate research which are never cited.

References in theses can include books and periodicals. In this study both were included. The total number of original references (excluding *ibids.*, *op. cit.*, and the like) per department were counted.

1. References in graduate theses (one year only).
2. References in graduate theses (five-year cumulation).

#### I. Other variables.

Other variables which might be in-

<sup>6</sup> William E. McGrath, "Measuring Classified Circulation According to Curriculum," *CRL*, XXIX (September 1968), 347-50.



cluded are reserve book use; citations in faculty publications; new periodicals published by subject; the total holdings in "complete" libraries, such as the Library of Congress, to be used as a comparison; books consulted in the library and left on tables.

The immediate objective of this study was to carry out a statistical analysis of the variables and to discover their relationships and relevance to each other and to the number and cost of books published annually. These relationships are interesting in themselves. The ultimate objective was to derive a simple formula which would describe the departmental book needs in relationship to books available and books used. In seeking this ultimate objective, the analysis reduces the number of variables, reduces the data to their simplest form, determines the best predictors, and, ideally, predicts the needs of academic departments in relation to each other.

## PART II

### ANALYSIS OF VARIABLES AND SPECIAL RELATIONSHIPS

In this study, *multiple regression* was first used, but it led to no special insights. A simple *multiple correlation* and the more sophisticated *factor analysis*, however, led to several insights.

#### *Multiple Correlations*

All forty-three variables listed in Part I were fed into the multiple correlation computer program, and a 43 x 43 correlation matrix was produced. This was done separately for each of three years. The three years correlated very highly with each other. Since data for the first year were somewhat sketchy, only the last two years were used in the final study; they were added together to ensure greater reliability.

Study of the inter-correlations permit-

ted the elimination of twenty-one of the forty-three variables from further study. Some of the twenty-one were dependent derivatives of others, which usually guaranteed a very high correlation. For example, some variables, such as averages of cost of books, credit hours, and contact hours, were all derivatives or correlatives of their totals. Others simply had insufficient or faulty data. For example, the detailed breakdown of circulation and inter-library loans by faculty, students, books, or periodicals in various combinations were unreliable because of the small numbers involved. A larger body of data on these variables would certainly justify a close study of them.

Table 2 gives a reduced matrix of correlations, using the twenty-two remaining variables. The correlations are on the Pearson scale. That is, a perfect correlation has a coefficient of 1.0; no correlation has a coefficient of 0.0; and a negative, or inverse, perfect correlation has a coefficient of -1.0.

The table shows a wealth of high correlations. Arbitrarily, anything above .70 is regarded as high. All coefficients, high or low, have a meaning of some kind. The variables with high correlations are useful in that one can be used to predict another. It is also useful to know that two with low correlations have little to do with each other.

Although the formula is not derived initially from these correlations, they do give considerable insight into the relationships of the variables to each other and can help to clarify their role in the factor analysis.

Consider the variables *Number and total cost of books published* (A-1 and A-2). These two have a very high correlation coefficient (.99), telling us that either number or cost gives us nearly identical percentages. This is enormously useful. If the percentage of books published is known, the percentage of cost is also known and vice versa—within a degree of accuracy, of course.

TABLE 2  
CORRELATION MATRIX

	A-1	A-2	B-1	C-1	C-2	C-4	C-6	C-8	D-1	D-2	D-3
A-1	1.00										
A-2	.99	1.00									
B-1	.93	.94	1.00								
C-1	.61	.65	.68	1.00							
C-2	.41	.44	.53	.84	1.00						
C-4	.73	.76	.78	.94	.79	1.00					
C-6	.53	.58	.63	.93	.85	.91	1.00				
C-8	.61	.66	.70	.95	.85	.98	.95	1.00			
D-1	.65	.70	.79	.84	.73	.89	.89	.90	1.00		
D-2	-.27	-.23	-.14	.24	.18	.22	.34	.27	.29	1.00	
D-3	.35	.39	.51	.75	.65	.77	.83	.80	.88	.71	1.00
D-4	.72	.76	.77	.93	.79	.99	.91	.98	.89	.24	.78
D-5	.60	.65	.71	.89	.75	.93	.94	.94	.96	.43	.93
E-1	-.17	-.14	-.10	.24	.33	.23	.35	.29	.36	.73	.63
E-2	-.25	-.20	-.01	.46	.40	.33	.51	.44	.49	.76	.74
E-3	-.16	-.13	-.10	.22	.31	.22	.33	.27	.35	.71	.61
F-2	.23	.27	.33	.58	.54	.61	.69	.64	.72	.75	.90
F-8	.94	.96	.98	.76	.57	.85	.70	.77	.80	-.10	.54
G-1	.19	.24	.20	.46	.26	.40	.59	.44	.58	.38	.61
G-10	.70	.74	.67	.62	.37	.67	.68	.64	.74	.00	.54
H-1	-.23	-.18	-.10	.02	-.03	.04	.13	.08	.23	.57	.44
H-2	-.18	-.14	.11	.04	.05	.04	.15	.11	.32	.43	.43
	D-4	D-5	E-1	E-2	E-3	F-2	F-8	G-1	G-10	H-1	H-2
A-1											
A-2											
B-1											
C-1											
C-2											
C-4											
C-6											
C-8											
D-1											
D-2											
D-3											
D-4	1.00										
D-5	.93	1.00									
E-1	.24	.45	1.00								
E-2	.33	.55	.62	1.00							
E-3	.23	.43	.99	.57	1.00						
F-2	.62	.80	.85	.66	.84	1.00					
F-8	.84	.76	-.06	.02	-.06	.37	1.00				
G-1	.41	.60	.33	.49	.31	.57	.26	1.00			
G-10	.68	.73	.05	.12	.05	.43	.72	.70	1.00		
H-1	.05	.24	.33	.50	.31	.38	-.10	.40	.03	1.00	
H-2	.04	.24	.09	.55	.06	.24	.03	.27	.02	.66	1.00

A-1 and A-2 also correlate highly with the *Existing collection* (B-1) and *Circulation by subject* (F-8). Their high coefficients (from .93 to .96) supports the ideas that (1) the subject output patterns of U.S. publishers does indeed reflect academic interest and has not changed much in recent generations,

and that (2) book use by department-as-subject conforms closely to what is available.

The low correlation (.37) between *Circulation by person* (F-2) and *Circulation by subject* (F-8) should dispel once and for all the myth that a person's department has much to do with the

subject of books he takes out. If the myth is true, then we must suspect that our classification systems (LC or DDC) fail to classify books properly, or that we should instead be classifying *persons* by subject. More likely, an individual's *specific* interest conforms loosely, if at all, to the interest of the *general* discipline he is teaching or studying. Librarians may be good examples of this. How many of the books that librarians read last year were actually books on library science?

*Number of faculty members* (C-1) correlates fairly (.76) with *Circulation by subject* (F-8), but not so fairly (.58) with *Circulation by person* (F-2). Since F-2 includes both faculty and students, it can be seen that the faculty does have some influence on the students—or is it vice versa?

Inter-library loans do not seem to reflect quite the same picture. There is a fair correlation (.70) between *I.L.L.'s by person* (G-1) and *I.L.L.'s by subject* (G-10). This relationship needs further study.

*Credit hours being taught—totals* (C-4) is an important variable, as shall be seen. High correlations between *Contact hours* (C-6) and *Equated hours* (C-8) have little meaning for us since those variables are functions of C-4. High correlations are also expected with other aspects of credit hours (D-1, D-3, D-4, and D-5). The importance of C-4 is its high correlation (.85) with *Circulation by subject* (F-8). Apparently, courses taken each semester do have a strong effect on the subject of the books taken.

Undergraduate variables (D-1, E-3) correlate better with *Circulation by person or major* (F-2) than do the graduate variables (D-2 and E-2). This is to be expected since graduate students account for a small portion of total enrollment and total circulation.

*Enrollment by declared major* (E-1,

E-2, E-3) correlates poorly with nearly everything except (1) *Credit hours offered or listed* (D-1, D-2, D-3, coefficients only fair, from .6 to .7), and (2) *Circulation by person* (F-2, coefficient .85 or .83).

The latter correlation is a clue to the greater relationships as revealed in the *factor analysis*. All the variables defined by person seem to be grouping together as do all those defined by subject—with little, if any, overlap between the two. Note that *Enrollment* (E-1) correlates well (.85) with *Circulation by borrower* (F-2), but poorly (-.06) with *Circulation by subject* (F-8).

Except for a modest correlation (about .5) with *Graduate credit hours and enrollment* (D-2 and E-2), *Citations in theses* (H-1 and H-2) have no high correlations with any other variables. The relationship seems obvious, since citations in theses are produced only by graduate students. The fact that they do not otherwise correlate with much of anything is significant. They seem to be independent variables. And therein is another clue to their significance which will be further revealed in the *factor analysis*.

#### General Significance

Coefficients need not be high to be significant, depending on what use we make of them. Two variables with a high coefficient means they both tell us the same thing, and we can discard one of them. Two with low coefficients indicate that one has nothing to do with the other, that one is not dependent upon the other, and that separately both are important. We must therefore account for both. In constructing a formula, we may have to use both.

The matrix shows many low coefficients. There are only a few negative, or inverse correlations, and these are all very low. None larger than -.27 shows up (between A-1 and D-2). We attach no special significance to this.



Many other combinations in the matrix could be discussed. Registrars and deans of students would be interested in the relationships between enrollment, credit hours, contact hours, and the like. Here we are interested mainly in the effect of these variables on use of the library, how they describe needs of departments, and how their inter-relationships can be used in a formula.

Without question these relationships will vary from institution to institution. Although some of them may be typical, we make no claim here that the findings are universal. It would be highly interesting and desirable to know which relationships are universal. This suggests the need for an inter-institutional cooperative study.

Some of the high correlations are not above suspicion. For example, the raw data for the Department of Languages and Social Sciences in variables A-1, A-2, B-1, and F-8 accounts for a very large part of the total, tending to overwhelm other departments and to promote high correlations among those variables.

Some individuals believe it is unfair to compare the humanities to engineering, or even the pure sciences to engineering. Others feel that experiments such as this are an excellent way to measure and compare actual differences. Whatever the plan, the investigator should consider the different relationships likely to result.

### *Factor Analysis*

Our formula is constructed from the results of a *factor analysis*, a device originally developed by psychologists for the study of personality.<sup>7</sup> Obviously

it can be used to study academic departments which, we might say, have corporate personalities. *Factor analysis* sorts out the complex relationships in the multiple correlations. We assume that if many variables can describe a person or a corporate body and that if some of these variables have something in common, the commonality can be discovered and precisely measured. When several variables overlap or group together (we have already seen this happening in the correlation matrix), these groups are called "factors." The analysis measures, on the Pearson scale, the precise amount of overlap.

The analysis will reveal as many factors as necessary to account for the desired amount of total variance. If twenty-two variables are used, the largest number of factors would be twenty-two and would account for 100 per cent of the variance. The object of the analysis, however, is to see if the number of factors can be reduced, with an acceptable amount of unaccounted variance. The investigator establishes the amount of variance he is willing to forego—say 10 per cent. The analysis will then produce the number of factors to meet this condition, say four. If we wanted to deal with only three factors, but this meant increasing the variance to, say, 40 per cent, we would prefer staying with four. Likewise, if decreasing the variance to 8 per cent meant an increase of five or ten more factors, again we would stay with four at 10 per cent. Figure 1 shows how, in a successful analysis, the variance levels off quickly after the first few factors.

In our analysis the factors were reduced to three with a total unaccounted variance of .15 and four with an unaccounted variance of .10. We decided to use three factors, as shown in Table 3. In each factor, each variable, to the extent indicated by the coefficient on the right, represents a measurement of the same thing. Any one variable can repre-

<sup>7</sup> An excellent account of factor analysis is given by Joseph R. Royce, "The Development of Factor Analysis," *Journal of General Psychology*, LVIII (April 1958), 139-164. Programs for this device, now standard, are included in the software package for many computers. A complete multiple correlation matrix is typically part of the printed output. Trained computation center personnel can run them.



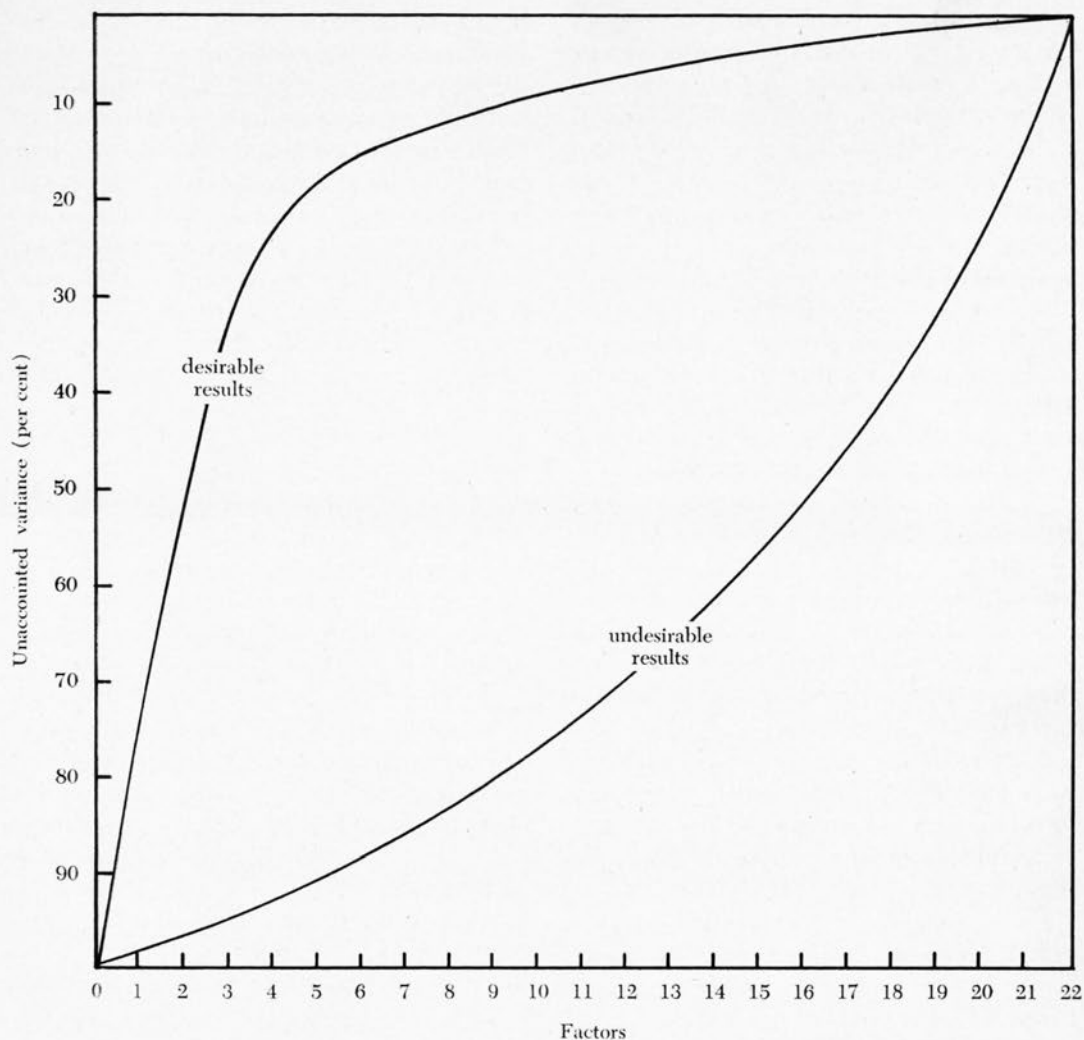


FIG. 1.—Hypothetical range of factors and variants.

TABLE 3  
THE FACTORS

Factor I		Factor II		Factor III	
Variable	Coefficient	Variable	Coefficient	Variable	Coefficient
A-1	.90	D-2	.77	G-1	.52
A-2	.92	D-3	.63	H-1	.82
B-1	.92	E-1	.95	H-2	.90
C-1	.87	E-2	.69		
C-2	.70	E-3	.94		
C-4	.93	F-2	.81		
C-6	.82				
C-8	.89				
D-1	.87				
D-4	.93				
D-5	.85				
F-8	.96				
G-10	.79				

sent that factor. The higher the coefficient, the better the representation. For example, in Factor II, E-1 is the best indicator.

Users of *factor analysis* play a little game called Naming the Factors. If factor analysis is truly a legitimate device and there are indeed factors, then, according to theory, they can be identified. More often than not, the players lose. They cannot identify the factors and must be content with simply numbering them. From the start of the project, before the analysis, it seems obvious to us (we hypothesized, you might say) that only three factors need describe departmental book requirements—*materials available or at hand*, *material used*, and *material not at hand and needed*. When the analysis gave us three factors at an acceptable level of variance, we were delighted because, surely, this bore out our hypothesis. But, try as we might, we could not make our three preselected names fit the three derived factors. Instead, it appears that the three derived factors should more appropriately be named I. *Subject of Books and Serials Used or Available*, II. *The Users*, and III. *Books and Serials Cited by Graduate Students in Theses*. All or most variables which somehow describe the subject of material used or available group together under Factor I. All or most variables which describe *the users* group together under Factor II. Even the names are not precise, and Factor III is something of a maverick. Until we have taken a closer look at these and other variables, we should be wiser to avoid names.

Our three "hypothesized" factors may still be valid; but if they are to have meaning, we must analyze other variables. Surely the factor, *material needed*, is valid; but none of our variables, with the possible exception of inter-library loans, seem to measure it.

If nothing else, this experience tells

us to test our assumptions and formulate our hypotheses carefully. If we are to be objective, we cannot let our wishes determine our conclusions.

Nevertheless, on the basis of data available and first-time statistical analysis of that data, we are justified in using what we have to derive a formula.

### PART III

#### THE FORMULA, STATISTICAL AND MATHEMATICAL BASIS

Since the *factor analysis* tells us that any one variable in its factor measures the same thing, to the extent indicated by its coefficient, we can use any one variable to represent the entire factor. This enormously simplifies the construction of a formula. Instead of using twenty-two variables in the formula, we use only three. In Factor I we have a wide choice of thirteen; in Factor II, six; and in Factor III, three.

As we said earlier, many of these variables have in one way or another been used by many libraries in arbitrary formulas. Each served its purpose after a fashion, but none of the libraries had any way of knowing whether the factors used were independent, non-repetitive, or even significant. We can now construct a formula which is more likely to consider the most significant and independent factors.

The actual formula used makes little difference as long as each factor is included. For simplicity, only one variable from each factor may be chosen, but two or more from each could be averaged. The criteria for selection should be (1) a high coefficient, (2) a substantial body of data, (3) easily collected data, and (4) resistance to deliberate local manipulation. A linear or geometric formula is a matter of choice. We chose the linear—*i.e.*, a simple additive formula.

For example, if F-8, E-1, and H-2 are chosen in Factors I, II, and III, we would add together, for each department, the fractional values or percentages for each of these variables. The allotment for one department is a fraction of the total amount to be divided. Because we have three factors, the fraction of the total is one-third the sum of the three fractions.

Basically, this is our formula:

$$\text{Factor I (F-8). } \frac{\text{no. of books taken out for that department}}{\text{total of all books taken out in all departments}} = \text{fraction of total (1)}$$

plus

$$\text{Factor II (E-1). } \frac{\text{enrollment in that department}}{\text{total enrollment in all departments}} = \text{fraction of total (2)}$$

plus

$$\text{Factor III (H-2). } \frac{\text{citations in theses from each department}}{\text{total citations in theses from all departments}} = \text{fraction of total (3)}$$

If the factors are all equally important, then the formula is fine as it stands. But if they are not, we must discover, somehow, which is more important, or arbitrarily decide which we *want* to be more important, and then weight them accordingly. Since discovery of an absolute weight is not within the mathematical capability of this technique, we must decide ourselves which is more important and assign the weight arbitrarily. We can assign weights by multiplying each of the three fractions by any number, as long as the three numbers add up to 1. After weighting each factor, for each department, the fraction of the total amount to be divided is the following sum:

$$\left[ \frac{\text{no. of books taken out for that department}}{\text{total of all books taken out in all departments}} \right] \text{ times } \left[ \text{weight of Factor I} \right] \quad (4)$$

plus

$$\left[ \frac{\text{enrollment in that department}}{\text{total enrollment in all departments}} \right] \text{ times } \left[ \text{weight of Factor II} \right] \quad (5)$$

plus

$$\left[ \frac{\text{citations in theses from that department}}{\text{total citations in theses from all departments}} \right] \text{ times } \left[ \text{weight of Factor III} \right] \quad (6)$$

Remember that

$$\left[ \text{weight of Factor I} \right] \text{ plus } \left[ \text{weight of Factor II} \right] \text{ plus } \left[ \text{weight of Factor III} \right] \quad (7)$$

must equal 1. If each of the three weights are .33, we have given the three factors equal weighting. We can give no additional advice on how much to weight each factor; this must be a judgment based on experience and the librarian's own knowledge of his own library. Mathematically, however, the fac-

tor analysis will provide a percentage figure for the amount of variance accounted for by each factor. One could use such figures, remembering that they represent an inherent weighting which may have nothing to do with the importance the librarian attributes to the factors.

Our formula is nearly complete. It lacks one important feature. In order to guarantee each department equality before the laws of the library, we might want to give each an equal amount to start. The amount could be nothing or it could be one hundred dollars or five hundred. The amount given, like the weighting, is arbitrary; and this part of the formula is not derived from the analysis. If we do allot an equal minimum to each department, the amount is

$$\frac{\text{amount to be divided equally}}{\text{number of departments}} \quad (8)$$

When we add (8) above to (4), (5), and (6), we have the final complete formula. The formula is the fraction to be multiplied by the total dollars to be divided. For those who want to read the formula in mathematical symbols, we have

$$A_D = \frac{E}{N} + (W_1 F_{1D} + W_2 F_{2D} + \dots + W_n F_{nD}) \cdot (T - E) \quad (9)$$

Where

- $A_D$  = Allotment for individual department (D)
- $E$  = Amount to be divided equally
- $N$  = Number of departments
- $F$  = Fraction of the variable (Factor) contributed by department (D)
- $T$  = Total amount to be divided
- $W$  = Arbitrary weighting value for each factor
- Sum of  $F_D = 1.00$

Bear in mind that we do not necessarily recommend use of the three variables mentioned, nor even that only three be used. Any three, or any other number can be used. The choice is entirely up to the librarian or his committee, and the choice is a function of his or their assessment of the data, its reliability, and the validity of the method.

As with any statistical device, its use here is to assist in a management decision. The statistics themselves cannot make this decision. ■ ■

