

South of Explosive Exponentialism in Academic Libraries

Being unfamiliar with significance levels, and correction techniques, librarians are overly concerned with lack of absolute precision in library counts. The handling of normal error is normal for modern statistics. From the admittedly imperfect statistics collected by university librarians since the 1940s, conclusions that are useful and statistically valid may be drawn. For example, the manifest "snowballing" trend in number of volumes in the nation's largest academic libraries greatly outweighs overall growth. Significant conclusions may be drawn also from geographical location and chronological development of the largest academic libraries.

SOME OF THE WISEST, wittiest, and most scientifically naïve literature imaginable has been written by librarians in their professional journals on the fine art of lying with statistics, on the frantic difficulties of counting and reporting library statistics accurately and consistently. Librarians seem to think that they invented these difficulties. Perhaps they have. There is no mention of significance levels, confidence limits, not to dream of the far-from-now-sophisticated matter of the chi-square test, and the student's t-distribution. If geologists can put into useful words and symbols what a fine-grained, light red, slightly rounded pebble is, so that another geologist ten thousand miles away will know precisely what he is reading, surely librarians can functionally define a "faculty member," or how many books occupy a foot of shelving. Well, what is a pebble, or a virus, or a book? A lot of things have happened in the world of statistics since

the days when the *Literary Digest* took a very large and inappropriate sample of people who owned telephones to predict the results of a presidential race.

Comparison of statistics of libraries of the South involves as background statistics of libraries of the entire United States. Moreover, comparison of any statistics is a great deal like working with grains of sand, not because the quantity of the data increases (this is handled easily with computers), but because the intriguing aspects, the almost innumerable possibilities of meaningful comparisons within the data grow more and more complex and beguiling. This quality is inherent, of course, in all mathematical studies; also, the more that is digested and interrelated in a mathematical approach to any data, the smaller and more concise becomes the statement of results. Furthermore, if the study is not directed toward mathematicians, it rapidly reaches a language barrier. Un-

fortunately, $\chi^2 = \sum_{i=1}^k \frac{(o_i - e_i)^2}{e_i}$ does not translate into adequate words.

Mrs. Hughes is at Texas Woman's University in Denton.

SCOPE OF SAMPLING

Sampling is one of the most important concepts in the study of statistics. It is basic to statistical study of libraries, as well as in any other field. The sample for the present study has been determined to be certain published statistics of the forty largest academic libraries in the United States as compiled by the U.S. Department of Health, Education, and Welfare, Office of Education (USOE) for 1964-65,¹ because first, data for these large academic libraries are more complete, of longer duration, and more comprehensive than for any other segment of the population of libraries in the South; and because second, one would not find a well-developed academic library existing in a library-vacuum. Does it not make sense to assume that good academic libraries breed in library-oriented communities and areas? Accepting this assumption, the phenomena of large academic libraries should be excellent sample observations of the trends in the whole library picture of an area.

Table 1 is reproduced from "Table E. Selected data for the 40 largest academic libraries: United States, 1964-65."² To this table have been added figures, lines, and symbols for the purpose of highlighting the libraries of the South against those of all other parts of the country. It is necessary to reduce this mass of data in the table to an understandable form: 1) for the analysis of data, which is the process of extracting from the measurements or counts relevant information from which a summarized numerical description can be formulated; and 2) for the interpretation of these data,

which is the process of drawing conclusions from the analysis of the data, and usually involves the formulation of predictions concerning a large collection of objects (or events) from the information available from a small collection of similar objects (the sample). These processes involve the application of statistical measures.

Most librarians who have been concerned with library statistics seem to overemphasize the detailed accuracy that they feel is essential if valid use is to be made of the comparisons. If in other sciences the world were waiting for absolute accuracy in sampling, as library science waits, man would not be circling the moon, and the course toward the discovery of the cure for cancer would await perfect counting and sampling techniques. The handling of normal "error" is normal for competent statistical study. As a matter of fact, statistics were developed because it is impossible to be entirely uniform and consistent, to count and record every item neatly and precisely. Statistics amount to the "best guess we can get, on the best information we can get." And that is considerably better than an opinion based on nothing. This is not to deny the desire for as much accuracy as we can reasonably get, and as workable, reasonable definitions as we can devise; it simply points to the fact that under the best of circumstances there will always be error and lack of uniformity, in every count, in every undertaking. To find something close to the truth in spite of error and lack of uniformity is why probability and statistics were recently invented and are such moving forces in the modern world. We can do a lot with what the library world has already collected in the way of counts and measurements. We are fortunate that librarians began seriously to count (however imperfectly) in the 1940s.

¹ U.S. Department of Health, Education, and Welfare, Office of Education, *Library Statistics of Colleges and Universities, 1963-64. Analytic Report* (Washington, D.C.: Government Printing Office, 1968), p. 7. This report includes "Table E. Selected data for the 40 largest academic libraries: United States, 1964-65," p. 7.

² *Ibid.*, p. 7.

TABLE 1

ADAPTED FROM: SELECTED DATA FOR THE 40 LARGEST
ACADEMIC LIBRARIES: UNITED STATES, 1964-65

Rank (1)	Institution (2)	Total Volumes ^a (3)	Volumes Added (4)	Total Library Operating Expenditures (5)	Books & Other Library Materials ^b (6)
← Mean	1. Harvard University	7,445,072	257,631	\$5,760,585	\$1,433,351
	2. Yale University	4,826,148 ^c	128,281	3,029,624	642,291
	3. University of Illinois	3,888,983	150,049	3,418,200	1,183,408
	4. Columbia University	3,569,565	123,311	3,173,973	997,960
	5. University of Michigan	3,376,076	135,533	3,552,508	897,109
	6. University of California—Berkeley	3,113,024	165,594	4,588,759	1,381,015
	7. Cornell University	2,725,624	152,822	3,281,869	1,047,226
	8. Stanford University	2,560,220	181,745	2,626,046	814,131
	9. University of Chicago	2,406,142	109,390	2,110,586	692,648
	10. University of Minnesota	2,405,797 ^c	88,896	2,160,561	726,999
← Median	11. University of California—Los Angeles	2,197,175	193,576	4,132,807	1,475,737
	12. University of Wisconsin	1,901,048	137,399	2,551,913	964,822
	13. University of Pennsylvania	1,894,480	89,880	1,824,686	580,125
	14. Indiana University	1,771,900		2,423,999	1,051,056
	15. Princeton University	1,769,699	65,934	1,597,881	595,666
	16. Ohio State University	1,748,943	88,173	2,019,516	580,732
	17. University of Texas ^d (1951-52) ^e	1,724,332	76,023	4,838,097	3,813,068
	18. Duke University ^d	1,716,855	70,465	1,417,431	560,578
	19. Northwestern University	1,709,172	68,796	1,438,691	496,735
	20. New York University	1,535,583	68,439	1,584,079	484,939
21. Johns Hopkins University ^d	1,399,700 ^c	76,342	1,005,272	306,295	
22. University of Washington	1,390,636	69,741	2,452,199	679,417	
23. University of North Carolina ^d	1,314,359	79,763	1,513,317	600,924	
24. Louisiana State University ^d	1,237,171	74,861	1,741,033	701,644	
25. University of Iowa	1,226,254	51,864	1,275,624	569,131	
26. University of Virginia ^d	1,221,353	65,798	865,983	329,241	
27. Rutgers University	1,207,350 ^c	72,781	1,508,467	610,587	
28. University of Missouri	1,167,000 ^c	58,777	1,240,910	523,793	
29. Michigan State University	1,147,188	97,660	1,460,133	587,218	
30. University of Southern California	1,138,812	47,177	1,169,192	339,335	
31. University of Kansas	1,122,158	54,417	1,206,368	485,975	
32. Wayne State University	1,116,620	65,490	1,552,434	546,093	
33. Brown University	1,111,240	34,197	1,007,934	313,391	
34. University of Florida ^d	1,087,665	52,268	1,418,317	480,646	
35. University of Kentucky ^d	1,069,908 ^c	38,345	1,060,677	419,015	
36. University of Oregon	1,057,879 ^c	79,835	1,056,789	343,825	
37. University of Oklahoma ^d	981,012 ^c	51,666	701,604	301,116	
38. Massachusetts Institute of Technology	959,212	64,367	1,003,942	240,649	
39. Joint University Libraries	947,941 ^c	40,370	800,736	289,442	
40. University of Colorado	946,435	70,119	1,451,715	451,212	

^a Institutions ranked by size of collection.^b Includes binding.^c Includes microform.^d Southern academic libraries.^e Passed the 1,000,000-volume mark.

Total Volumes = 77,135,731

X = 2,075,611.4 books

s = 1,194,840.9 books

Range = 6,498,637 books

ANALYSIS OF SAMPLE

The forty largest academic libraries listed in Table 1 are arranged in descending ranked order according to total volumes (column 3). The mean of the distribution is 2,075,611.4 volumes; the median is 1,600,000 volumes; the range is 6,498,637 volumes; and the standard deviation is 1,194,840.9 volumes. The analysis of the data was accomplished on a computer, and resulted in forty punched cards, and a Program Sheet.

INTERPRETATION OF SAMPLE:
GEOGRAPHICAL

Consider in Table 1 the forty libraries by geographical area: Northeast, North Central, South, and West. With ten of the largest academic libraries in the South, compared to ten in the Northeast, thirteen in the North Central, and seven in the West, superficially it would appear that the South is very well off. That is far from true. When we see these libraries plotted on a map (Figure 1),³ it is obvious immediately that there is a tremendous difference in density of great libraries by area. And when we compare this outline map with an ordinary road map of the coterminous United States, one is immediately impressed with the vast differences in the human geography of the regions—in communication lines, in population centers, in the whole network of inhabitation. The ten libraries in the South cover very poorly their large area of land and their share of the population.

CHRONOLOGICAL DEVELOPMENT

Table 2 is a graphic presentation of some of the same forty libraries arranged by rank again (number of volumes) and in column 2 of this table is given the age of all the institutions supporting the

largest libraries in the South compared to the ages of all the institutions supporting the great libraries *above* the largest Southern library (the Library of the University of Texas). Here we see that the academic environments supporting the largest libraries range in age from Harvard's 332 years to the University of Oklahoma's 78 years (excepting the composite age of the Joint University Libraries). The University of Texas was only 87 years of age in 1965, but so was the University of California in Los Angeles, which has far outranked it in acquisitions. Why? Surely not wealth. Both areas have long been affluent. The significant trend here seems to be population. Libraries grow most where most of the people live; libraries grow most rapidly in the Standard Metropolitan Statistical Areas. Many of the Southern universities in the group are as old or older than academic institutions in the advantaged group. The University of Virginia was 149 when Stanford was only 83. Age is seemingly not a dominant factor. The density of the largest libraries follows the density of the population, the SMSA's.

In the major body of Table 2, it is shown by means of rectangles (hatched and stippled) when each of these selected largest academic libraries passed the 1,000,000-volume mark. Texas made it in 1951; Duke at practically the same time. All six of the top largest libraries were already past the 1,000,000-volume mark *before* 1941. Two of the next ranking group, Chicago and Minnesota, were over the mark in 1941. *By* 1954 all of the leaders had at least 1,000,000 volumes. Significantly, six of the largest Southern academic libraries passed the point only six years ago. Obviously, there has been considerable delay in library growth in the academic libraries in the South.

Returning to Table 1, consider the choice of data. The six columns given in Table E from which these data were chosen present values for: Total vol-

³ The area divisions are reproduced from an outline map of the U.S. Department of Commerce, Bureau of the Census, published in the *U.S. Book of Facts, Statistics & Information*, 88th ed. (New York: Essandess Special Edition, 1967), xiv.

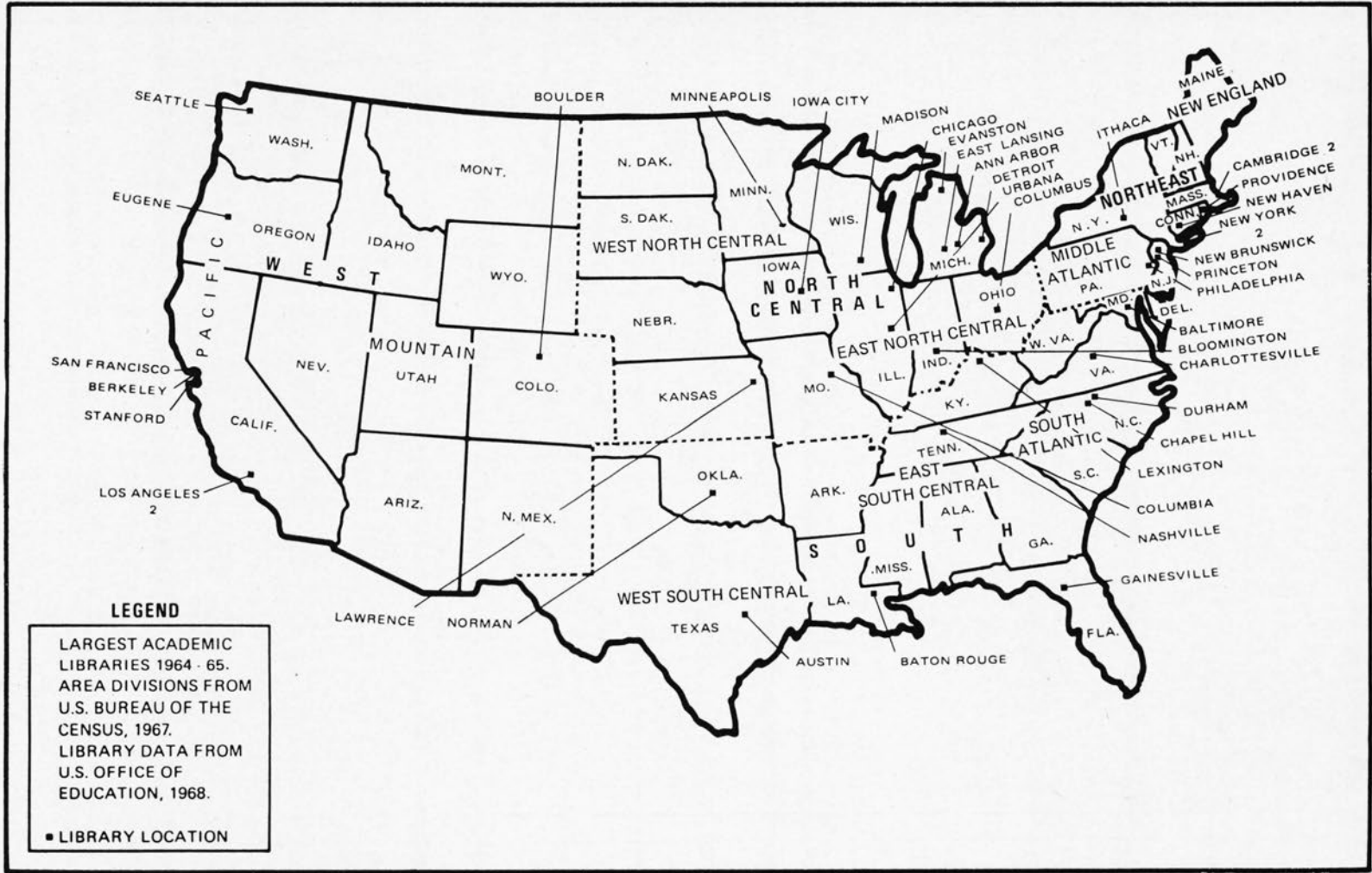


FIGURE 1.—LARGEST ACADEMIC LIBRARIES 1964-65

	AGE	1941 or-	1946	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956	1962 or later+
1. HARVARD UNIVERSITY	332	■												
2. YALE UNIVERSITY	267	■												
3. UNIVERSITY OF ILLINOIS	100	■												
4. COLUMBIA UNIVERSITY	214	■												
5. UNIVERSITY OF MICHIGAN	151	■												
6. UNIV. OF CALIF. BERKELEY	100	■												
7. CORNELL UNIVERSITY	103				■									
8. STANFORD UNIVERSITY	83					■								
9. UNIVERSITY OF CHICAGO	78	■												
10. UNIVERSITY OF MINNESOTA	117	■												
11. UNIV. OF CALIF. L.A.	87										■			
12. UNIVERSITY OF WISCONSIN	120										■			
13. UNIVERSITY OF PENNSYLVANIA	228		■											
14. INDIANA UNIVERSITY	148										■			
15. PRINCETON UNIVERSITY	222										■			
16. OHIO STATE UNIVERSITY	98								■					
17. UNIVERSITY OF TEXAS	87									■				
18. DUKE UNIVERSITY	130						■							
21. JOHNS HOPKINS UNIVERSITY	92												■	
23. UNIVERSITY OF N. CAROLINA	179													■
24. LOUISIANA STATE UNIVERSITY	108													■
26. UNIVERSITY OF VIRGINIA	149													■
34. UNIVERSITY OF FLORIDA	115													■
35. UNIVERSITY OF KENTUCKY	103													■
37. UNIVERSITY OF OKLAHOMA	78													■
38. MASSACHUSETTS OF TECHNOLOGY	107													■
39. JOINT UNIVERSITY LIBRARIES	32													■

LEGEND:





 SOUTHERN	 NORTH CENTRAL
 NORTHEASTERN	 WESTERN

TABLE 2.—CHRONOLOGY: MORE THAN 1,000,000 BOOKS

umes; Volumes added; Total staff in full-time equivalents (FTE); Total library operating expenditures; Staff salaries and wages; Expenditures for books and library materials (including binding); and nine of the libraries include microform in their counts of total volumes. Much has been written on the choice and potential use of these and other similar categories. For example, John Weatherford wisely warned against using the figure for numbers of faculty in an institution to derive the ratio of library expenditure per faculty member because a "good" ratio can be obtained by having a small faculty as well as by having a large library budget.⁴ In Radford's excellent study of the problems of academic library statistics⁵ he has made many important observations and contributions (as well as a number of astute, but rather mathematically "picayunish" comments). Table 1 does not indicate whether or not "Volumes Added" is net. It would be useful to know. It is also progressive when Radford notes that the three major statistical surveys of academic libraries (USOE, ALA, and ARL) have all "concentrated their efforts on the expenditures of libraries, ignoring the income side of library finance."⁶ The array of forty libraries in Table 1 shows that Harvard is Number One in every category except expenditures for "Books & other Library Materials." There it is outranked by the library of the University of Texas. However, the ALA 1965-66 report⁷ placed the University of Texas library expenditures at \$1,070,083, a more valid figure. Number 40 in the volume-ranked list in Table 1 is the University of Colorado, but this institution is

far from the bottom in the important category of "Volumes Added."

Harvard is the giant, and increasing by some 257,000 volumes each year. It is impressive. There are only *seven* states (excluding Massachusetts), California, Illinois, Michigan, New York, Ohio, Pennsylvania, and Texas, in which the holdings of *all* the academic libraries, are larger than those of Harvard University alone. Of the top ten academic libraries, Harvard, Yale, Illinois, Columbia, Michigan, California at Berkeley, Cornell, Stanford, Chicago, and Minnesota, there is not one in the South.

Six of the top ten academic libraries are private institutions. But of the entire list, only fourteen of the forty largest academic libraries are private. In the past this has been a much larger percentage. There is a trend toward larger public academic libraries. Two of the largest academic libraries, Harvard and Yale, with more than twelve million of the overall total of some seventy-seven million volumes, are almost exclusively for the education of men.

What is the significance of the numerical values that were obtained in the analysis? What is the importance of knowing that in this array the mean is 2,075,611.4 volumes, and lies, therefore, just below the tenth item, the University of Minnesota Library; while the median, which can be ascertained at a glance (inasmuch as the array is in ranked sequence, the median is the middle of the forty items), and lies between the nineteenth and twentieth largest academic libraries—exactly between Northwestern and New York Universities? If this distribution fitted a normal curve, the mean and median would nearly coincide. Instead, we can observe from the lines drawn on Table 1 representing the mean and the median, that this distribution is definitely "lopsided," or strongly skewed toward the higher values. In other words, the variates (volumes in this case) plotted

⁴ John Weatherford, "USUAL: A Visit to the United States University Average Library," *Library Journal*, XC (December 15, 1965), 5345.

⁵ Neil A. Radford, "The Problems of Academic Statistics," *Library Quarterly*, XXXVIII (July 1968), 231-248.

⁶ *Ibid.*, p. 243.

⁷ American Library Association, *Library Statistics of Colleges and Universities, 1965-66. Institutional Data* (Chicago: American Library Association, 1967), p. 154.

graphically would not make the smoothly rising and falling slope, expected in a normal curve. Instead, we have a jagged picture. Most of the volumes, many more than normally expected, are in the few very largest libraries. For example, Harvard has a variance from the mean of 4,174,621 volumes, while Yale's variance is only 555,697 volumes. The University of Texas library looms as an abnormal peak in its position as number 17, with a variance of 843,561 volumes, while the University of Oklahoma library exceeds its expected value by only 100,241 volumes. The trend is a centralization movement, an exponential "snowball effect." The biggest libraries are getting bigger faster than the others. The total number of volumes in the aggregate United States college and university libraries in 1964 was 227,100,000 volumes, and of these, over 77,000,000 were in 40 out of 2,140 academic libraries.

Good students are attracted to good faculties; and good faculties are attracted to institutions having the best libraries. "And—and this is the whole point—libraries which are good, for the diversity of interests which are represented in a university faculty, necessarily are, or soon become, large libraries."⁸

Intertwined also (for better or for worse) are large, growing academic li-

braries, and large, spreading population vortexes. The most reassuring development in urban affairs in the past several years is the increasing awareness of the importance of the *quality* of the environment in which man lives. In 1967 the Department of Housing and Urban Development introduced a guide to cities applying for a grant in its Demonstration Cities Program. Applications for grants have been filed from over two hundred cities and counties. These planners must take into consideration the fact that man develops and demands great academic libraries in urban centers. Libraries exist because of man, and the social organizations he invents. Investigations of libraries are part of the comprehensive attack on human problems, and must include the methodology of modern statistics.

While correcting the flaws in our gathering of library statistics, it is advisable to remember that a 3 per cent margin of error is allowed in matters of life and death; and organizations risk capital, regularly and successfully, on a 5 per cent chance of error. Even considering the "family joke" of the library that jumped (in published statistics) from 50,000 volumes in 1962 to 100,000 volumes in 1964, by adding 5,000 volumes, we still have a sample that is worthy of statistical analysis. The situation could rapidly improve with enforced standards of what should be counted. ■■

⁸ Verner Clapp, "Graduate Education and Library Resources," *Journal of the Graduate Research Center of Southern Methodist University*, XXX (1962), p. 51.