6.0 : Objectives

After reading this unit, you will be able to:

- understand Bradford's law which describes how literature in a subject is scattered or distributed unevenly in journals;
- see how this uneven distribution can be used practically as an efficient cost-saving factor;
- identify the important of `core' journals in a subject;
- use standard data sets to generate the Bradford curve;
- generate data for decision making; and
- learn about standard and new applications of the Law of scattering.

6.1 : Introduction

In this unit we shall look at a very useful phenomenon observed in `bibliometrics' that is the analysis of published literature. It relates to the *'scattering' of* relevant articles in journals. What do we mean by *scattering? If* we wish to create a bibliography, or a complete list *of* published articles on a given subject, *say* Polymer Science, we know that we will have to scan through several journals because the articles are, of course, *scattered* or dispersed over many journals.

If we knew the nature of this scatter, i.e. where we are more likely to find articles, the search would become much easier for us. A mathematical relation describing dispersion of articles among journals was first stated by S.C. Bradford, Librarian in the National Science Library in London around 1934. In the rest of this unit we will discuss this relationship, now known as Bradford's law, and examine implications, usefulness of this law in a number of practical application in the management of library collections and abstracting services, as well as used increasingly in internet applications.

6.2 : Historical Account

Samuel Clement Bradford was chief librarian in the National Science Library in England from 1925-1938. While in this position he became concerned with the problems of omission and overlap in the coverage of primary journal literature by indexing and abstracting services. In an article published in 1934, he wrote:

"Those who are concerned with progress in science and invention are aware of the need for the provision of an efficient service for abstracting and indexing scientific and technical literature. It is therefore somewhat disquieting to find on inquiry that although the 300 abstracting and indexing journals notice 750,000 articles each year which is the same as the total number of papers published in their fields; owing t duplication of effort, only 250,000 different articles are dealt with and 500,000 are missed. It seemed worthwhile to inquire whether the cause of this failure might not lie in the manner in which the literature of a subject is distributed among the periodicals that contain it, ...the results fully confirm this view."

Bradford started from the premise that the problem was due to the overlapping ; nature of scientific
specialties, stating every scientific subject is related, more or less remotely, to every other scientific *subject,..* and that therefore, .. the articles of interest to a specialist must occur not only in the periodicals specializing on his subject, but also, from time to time, in other periodicals.."

He applied this principle in an analysis of two specific subjects by examining bibliographies, or lists of papers, compiled for Geophysics and Lubrication. He then prepared lists of journals; where the journals were arranged by decreasing order ~f the number of articles contributed by them to the bibliographies. He noticed that in each subject there were very few productive journals that contributed many articles to the bibliography. A large number of journals that contributed a few articles each and an even larger number that contributed one or two articles to the bibliography This empirical observation formed the basis of a mathematical relationship regard **€** the number of journals of different levels of productivity. The relationship is n known as Bradford's law. The results were published in a paper in 1934 and i book called Documentation in 1948.

Bradford's work became influential because of it applicability in efficient retrieval of information. His book brought together his research that laid the groundwork 'o the many applications for citation analysis and bibliometrics. It now appears to concern about issues on the management of libraries existed since ancient tit e: The origins of bibliometric studies have been traced back to the ancient Egyptian Alexandrian Library in a study by Nisonger. In India, Ranganathan was a pioneer in the use of statistical methods in the management of libraries. Although such studies have occurred down through the centuries, Bradford's Law formulate 1934 again brought bibliometric studies into prominence.

6.3 : Bradford's law

Let us consider the number of papers in a given subject that we may expect to get if we search in a journal library. By scanning through sufficient journals we expect to find most of the relevant articles. We have some ideas which are the most relevant journals where we are most likely to find articles we need. For example specialist journals in the given subject area will have more articles that will suite our need. In addition, there would be some articles in general journals which cover other areas as well. There would also be an occasional article in some journals, while most journals would yield nothing.

Clearly, we should not do a random search but prioritize our search so as to the maximum number of journals in the least time. First we will search through the pages of technical journals in the subject because the likelihood of finding useful papers is higher than in more general journals. After we have looked at these journals, we may wish to extend our search to more general journals covering our area of interest. Finally, we may move on to journals which report on all fields, in which we may still find a-few papers of interest. At the end of this exercise, we will have a complete list of papers or `bibliography' in the subject on which we had set out to conduct the search.

On the other hand, we may decide to end the search once we have searched through the specialized journals, as it becomes more time consuming and costly to search the very large number of remaining journals. At this point we do not know how many journals of each kind there may be and how time consuming or expensive the search may turn out to be. This is where Bradford made his important contribution and to do this, he began by ranking the journals in order of the number of articles found in

them.

The intuitive procedure for searching described above can be facilitated by ranking the journals in order of their `productivity', i.e., how many relevant articles they contain. `Bradford arranged all the journals in the bibliography in descending order of productivity. He took the journal which contained the largest number of relevant articles and placed it at the top of a list and gave it the 1st rank. The next most productive journal was given rank 2, and so on. If two journals had the same rank R, then the next journal would have rank R+2. *[Note: The rank R of any journal indicates that there are R-1 journals which have higher productivity.]*

Having obtained a ranked list of journals arranged in decreasing order of productivity, Bradford divided the list of journals into three parts or `zones', each containing approximately the same number of articles. The first zone contained the most productive journals, the next zone had less productive journals and the third had the least productive journals. Clearly, in order to yield the same number of articles would require many more journals of low productivity as compared to journals of high productivity. This means that the first zone, sometimes called the core or nucleus, would contain fewer journals but yield the same amount of information in terms of number of articles; these would also be the journals specializing in the subject. This, then, was the observation made by Bradford in 1934 about the number of journals in each zone, and he expressed his law of scattering in the following way:

"If scientific journals are arranged in order of decreasing productivity of articles on a given subject, they may be divided into a nucleus of periodicals more particularly devoted to the subject and several other groups or zones containing the same number of articles as the nucleus when the number of periodicals in the nucleus-and succeeding zones will be as 1: n: n^2 (i.e. 1:5:25)

According to Garfield, a physical analogy of the situation described by Bradford would be a cornet, with the dense nucleus representing the core journals of a field and the sparse tail representing additional journals that sometimes publish material relevant to the field; these are sometimes called the peripheral journals.

6.4 : Verbal Form of Bradford's Law

Bradford found that in any given subject the number of journals in each zone in decreasing order of productivity increases in a geometric progression, starting with the most productive journals j0, as j0, j0 r, j0 r2, where j0 is the number of core journals. The core journals are usually the most important journals for the subject. The parameter `r' is a constant for a given bibliography, also known as the Bradford multiplier.

This is known as the verbal form of Bradford's Law, since Bradford also gave another version of his relationship which will be described in the next section.

Bradford formulated his law after studying a bibliography of geophysics, covering 1332 articles from 326 journals in the field. He discovered that 9 journals contained 429 articles, 59 contained 499 articles, and 258 contained 404 articles. So it took 9 journals to contribute one-third of the articles, 9, or 45, to
produce the next third, and 5 times 9, or 225, to produce the last third. As may be seen, Bradford's Law is not statistically accurate, strictly speaking. But it is still commonly used as a general rule of thumb. It was later found by Vickery that the number of zones need not be restricted to three. In fact if a bibliography contains A articles contributed by a number of journals J, it can be sub-divided into a number of zones n, each contributing the same number of articles A/n, and containing j_0, J_1 , j_2 , j_{n-1}
$_1$ journals, then

$$j_1 = j_0 r$$

$$j_2 = j_0 r^2$$

$$\ldots\ldots,$$

$$j_{n-1} = j_0 r^{n-1}$$

In an example in the next section, we shall demonstrate how the same bibliography can be divided into different number of zones. Note, however that the value of the Bradford multiplier will depend on the number of zones and is not fixed for a given bibliography.

Bradford's law is not a law in the sense that it has not been proved and is not exact, but has been verified to be approximately true in many cases as to be sufficiently general and usable for most applications. We shall discuss the deviations from this law in a later section.

6.5 : Applications of Bradford's Law

Notice that each succeeding Bradford zone contains successively larger and larger number of journals. This means that if we start our search with the most productive journals, then as we go along, to retrieve the same number of articles, we have to scan more and more journals. In other words, there is a law of diminishing returns for the effort expended in conducting the search.

Application to abstracting services: Abstracting services would like to give as complete coverage as possible to the literature. However, the cost of scanning journals increases in proportion to the number of journals covered. On the other hand, the number of articles actually retrieved depends on the type of journal scanned. It is, therefore, good procedure to rank the journals as described in the Bradford sense and begin scanning from the most productive journal, followed by other core journals. In this way, one would get the maximum coverage at the lowest cost. Since the retrieved articles decrease as one proceeds from the most productive journals in the core to the less productive peripheral journals, whereas the cost per journal scanned remains constant, at some point, it is no longer cost effective to continue the search. This should become clearer once you go through the example given later in this section.

If an abstracting service wishes to provide *a comprehensive* bibliography, i.e. include all the papers published, it would be necessary to scan all journals. The *number of* journals to be scanned, and therefore the effort and time required to retrieve an equivalent *number of* papers would increase as it moved from the zone of highly productive journals to the less and less productive ones. This would imply increasing costs. At some point along the way, the cost may exceed the money that the abstracting service expects to retrieve from the persons who would subscribe to the service. Such a service would then not be economically viable and would have to be stopped

On the other hand, because of the concentration of articles in the most productive journals, by simply scanning say the core journals, the service covers one-third *of* the articles by searching through a much smaller fraction of the total journals.

Example 1: A complete bibliography of Flying Saucers of 3000 articles with 3 zones has a core size of 10 journals and a Bradford multiplier of 2. The cost of scanning a journal is Rs. 10, while the company provides articles to subscribers at Re 0.01 per article. If there are 200 subscribers, calculate what is the rate of profit for covering (i) the first zone, (ii) the first 2 zones, (iii) all the three zones.

Solution: Each of the 3 zones contain 1/3 of the total bibliography, e.g. 1000 articles. The first zone contains 10 journals. The next zone contains 10*2=20 journals. The 3rd zone contains 10*2^2=40 journals. Based on this we can construct the following table:

Zone	**1 (core)**	**2**	**3**
Articles	1000	1000	1000
Journals	10	20	40
Cost *of* scanning	Rs. 500	Rs. 1000	Rs. 2000
@ Rs. 50 per journal Cost to retrieve one article	Re. 0.50	Re. 1	Rs.2
Earning from 100 subscribers @ Re 0.01 per article	Rs. 1000	Rs, 1000	Rs. 1000
Profit	Rs. 500	Rs.0	Rs. 1000

i. Article coverage is 33% by scanning 10/70 = 14% journals at a profit rate of (500* 100/500) = 100%
ii. Article coverage is 66% by scanning 30/70 = 43% journals at a profit rate of 500* 100/1500=33% iii. Article coverage is 100% by scanning 100% journals at *a loss* of 500* 100/3500 = 14%

This is an application of Bradford's law which indicates the loss of information when the search is limited by considerations of cost, etc.

Application to library holdings: Possible applications of Bradford's law could be to the question of journal subscriptions or planning space for a library. An optimal library size may be determined based on the scattering of articles among the journals in different subjects. If a library covers several subjects, due to their overlap the core journals in a particular subject will often serve as the peripheral journals for other subjects. The overlapping nature of disciplines implies that the core journals in a field will some times publish articles in adjoining disciplines. In other words, they act as peripheral journals for other fields. Thus, the core journals for a set of subjects will automatically include some of the peripheral journals in these subjects ensuring better coverage all round. Library size does not increase in proportion to the number of areas covered, and a general science library covering 10 fields is not necessarily 10 times larger than a specialized library covering a single field. This has important implications for resource allocation of space and finance in the planning *of* libraries and indexing services, because it states that much of the scientific literature is concentrated in a set of journals that form the core of the scientific literature.

Garfield's Law of Concentration: The extension of Bradford's law from one to several disciplines has been put in the form of a law by Eugene Garfield, a pioneer in the field of information. Garfield was publishing several indexes, including the Science Citation Index (SCI), through the Institute for Scientific Information (ISI) in America and processing citations to journal articles in several disciplines. His studies revealed that a multidisciplinary mix *of* 152 journals accounted for 50% of the citations processed for the SCI in 1969. This finding caused Garfield to formulate his law of concentration, which he derived from Bradford's law of scattering by extending it from the level of a single discipline to that of science as a whole

Garfield devised a physical analogy to Bradford's law, one based upon a cornet, where, the nucleus of the comet represents the core journals of a discipline's literature with the debris and gas molecules of the tail representing the additional journals that sometimes publish material relevant to the discipline. With this analogy in mind, Garfield described his law of concentration and its practical implications in the following manner

" .... The bibliographic law of concentration goes a step beyond the Bradford law by stating that the tail of the literature of any one discipline consists, in large part, of the cores of the literature of all other disciplines, and that all the disciplines combined produce a multidisciplinary literature core for all of science that consists of no more than 1,000 journals. In fact, this multidisciplinary core might be as small as 500 journals. Though larger collections certainly can be justified in many cases, the single function of providing reasonably cost-effective coverage of the literature most used by research scientists require no more than 500 to 1,000 journals."

According to Garfield, Bradford's principle has been kept in mind in the provision of indexing services by ISI

6.6 : Graphical Representation of Bradford's Law; Core,

$$\ln \frac{R_2}{R_1} = \ln \frac{R_3}{R_2} = \ln \frac{R_4}{R_3} ......etc$$

To illustrate his law, Bradford also plotted the cumulative number of articles against the logarithm of the journal ranks. The cumulative number of articles Corresponding to the journal rank R is the total number of articles obtained from journals up to am including the journal of rank R.

The J shaped graph that Bradford obtained, also called the Bradford bibliography appeared as an upward curve before it became linear. The curve plotted for the bibliography of Lubrication used by him is shown in Fig.1.

Notice that the Bradford curve has a rising portion, consisting of journals with the highest productivity. The area described by the curving line is usually regarded the "nuclear zone" or journal core. This is followed by an almost linear portion or mid-section where the cumulative number of articles contributed by journals up to a given rank increases proportionately with the logarithm of the rank. This is also the part which corresponds to the verbal form of Bradford Law as we shall see below.

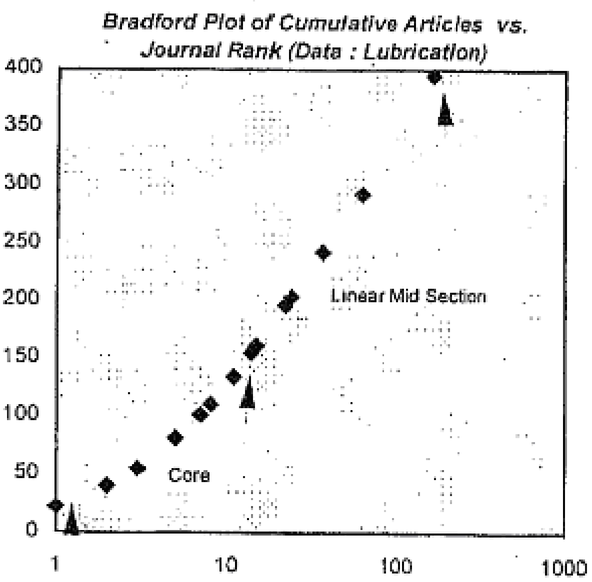


Fig. 1

If we divide the straight line portion into zones each containing the same number of articles, then each zone corresponds to a set of journals. Let the zone containing journals between ranks R_2 and R_I yield articles A_Z-A_1, the next zone containing journals between ranks R_3 and R_2 yield articles A_3-A_2, etc.

Since the slope of the straight line portion is constant, it is easy to see that

$$\frac{A_2 - A_1}{\text{In}R_2 - \text{In}R_1} = \frac{A_3 - A_2}{\text{In}R_3 - \text{In}R_2}$$

Also, since the zones contain the same number of articles Therefore,

$\text{In } R_2 - \text{In}R_1 = \text{In}R_3 - \text{In } R_2$

or,

$$\text{In}\frac{R_2}{R_1} = \text{In}\frac{R_3}{R_2} = \text{In}\frac{R_4}{R_3} \text{......etc}$$

The solution to this equation is a geometric progression.

$R_1 = a$
$R_2 = ar$
$R_3 = ar_2$

where a = the number of journals in the first zone and the number of journals in each successive zone of lower productivity increases by a factor of r.

Note that this is almost the same as the verbal form of Bradford's Law.

Bradford did not give a mathematical model for his law. Models were suggested later by Brookes, Vickery and Leimkuhler.

6.7 : Conditions for Bradford's Law

It was suggested by Brookes that for strict conformity with Bradford's law certain conditions have to be imposed on the bibliography. They are:

- The subject of the bibliography must be well-defined
- The bibliography must be complete; that is, all relevant papers and serials must be listed,
- The bibliography must be of limited time-span so that all contributing serials have the same opportunity of contributing papers.

However, it has also been shown by Bookstein that the shape of the curve is fairly robust and does not change much in spite of the variables such *as* subject, time span etc. not being very well defined.

6.8 : Falling Tail of Bradford's Curve ; The Groos Droop

In 1967, it was noticed by Groos that the curve plotted from actual data on journals and articles has a linear portion that falls off at the tail end of the less productive journals. This phenomenon is today known as Groos droop because, the curve drops at its end, giving it an S rather than a J shape, as seen in the bibliography of Agriculture plotted by Lawani, shown in Fig. 2. The sloping part at the top of the Bradford curve is called the **Groos droop.**

Data on Publications in Agriculture
(Lawani' *9373)*

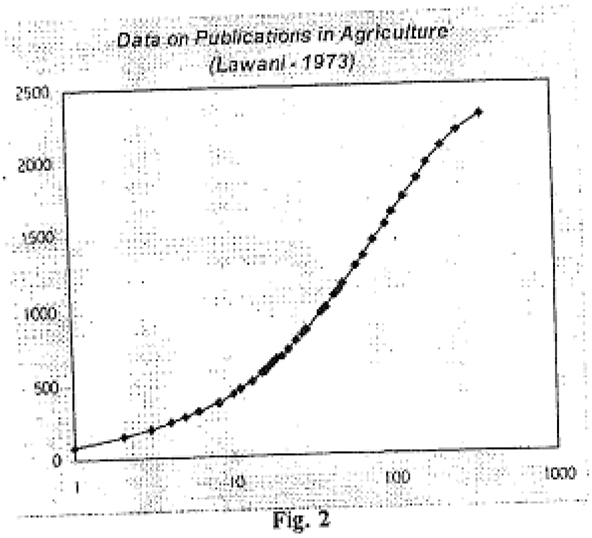


Fig. 2

Brookes discussed the Groos droop and argued that the droop was an indication 3f the incomplete nature of the bibliography examined. His hypothesis was that to strength of the Groos droop reflects the (in) completeness of a bibliography. He also developed a method for estimating the total size of the bibliography.

However, opinion on the reason for the droop was divided. While Brookes felt at this was due to the incompleteness of the bibliography, others like Wilkinson and O'neill showed that in many cases the S-shaped curve gave a better description of the data plot than the J shaped curve.

In any case, this portion of the curve corresponds to the least productive journals, either non-specialized or specializing in other fields and therefore is not very important from a practical point of view. However, if one considers the theoretical basis of the law, as to what factors give rise to this pattern of scatter, then the details of shape become important.

6.9 : Ambiguity in Bradford's Law

Bradford's verbal formulation of the law of scattering was not mathematically equivalent to the graphical representation described in his original article in 1934. The verbal form was the result of theoretical speculation while the graphical form was obtained from empirical data derived from the two bibliographies with which he worked. The verbal portion corresponds to the linear portion of the bibliograph, but not to the entire graph with its rising portion. This was noticed by Vickery who also pointed out that the relation holds for any number of zones and not just three zones as in Bradford's original formulation.

The graphical formulation has been found to be more accurate when compared with data and also more convenient to use. The nucleus is easily identified and deviations from the verbal form of Bradford's law such as the Groos droop may easily be detected in actual observed data

6.10 : Fitting Bibliographic Data to Bradford's Law

Following Egghe and Rousseau we apply the fitting methods to the bibliography of

Lubrication, the data for which can be found in Bradford (1934). In principle, the number of zones n may be chosen freely, although some limitation is necessary since the data are finite. We first consider n = 3 and next n = 7 just to show that n can be chosen more of less freely.
Each zone contains a0 articles, where

AO=A/N [I]

r = *(ey* am)l/N = (1.781 am)I/N [2]

Further more, since j0 + jOr + ... + jOrN-l = J, by taking the sum of the geometric progression we get,

JO = J(r-l)

rN - 1

Table 1: Data of 'Lubrication (1934)

No. of journals(j)	Corresponding no. of articles (a)	Articles (col.1 *col.2)	Rank (R) of journal	Cumulative no. of articles
1	22	22	1	22
1	18	18	2	40
1	15	15	3	55
2	13	26	5	81
2	10	20	7	101
1	9	9	8	110
3	8	24	11	134
3	7	21	14	155
1	6	6	15	161
7	5	35	22	196
2	4	8	24	204
13	3	39	37	243
25	2	50	62	293
102	1	102	164	395

Table cont...

[Note on calculation of rank and cumulative number of articles:

- *The rank corresponding to any row is obtained by adding the number of journals with higher or equal productivity.*

- *The papers are obtained by taking the product of the first two columns. The cumulative papers corresponding to a given rank are the total of all papers obtained from journals of higher productivity.*

Applying it to the data of Lubrication, (table 1) we get the following:

1) Bradford data fit for number of zones n = 3

We have

r = (1.781 * 22)1/3=3:40

a0 = 395/3 = 131.67, 132 and

j0 = 164(r-l)/r3-1 =1.0.30. Hence we use [j0] = 10.

The zones for n = 3 are shown in Table 2,

Table 2: Lubrication data divided into 3 zones

	No. of journals	No. of articles	r
1st zone	j0= 10.30 _ 10	126	-
2nd zone	j0r = 35.02 _ 35	133	3.50
3rd zone	$J0r^2$= 119, which is exactly the last	136	3.40
	rank in the bibliography		

This is a better fit than Bradford's original example (Bradford (1934)) where he gets 8/29/127 journals yielding respectively 110/133/152 articles.

Exercise: Break the data as Bradford would have done and determine the vah of the core journals and the Bradford multiplier as obtained by Bradford.

1) Bradford data fit for number of zones n = 7

Here we find

$r = 1.69$, $a_0 = 56$ and $j_0 = 2.95 = 132$.

The Bradford zones: for N = 7 are shown in Table 3.

Table 3: Bradford zones for 7 zones (Lubrication)

	No. of journals	No. of articles	r
1" zone	J_o 2.95 =3	55	
2nd zone	$J_o r$ = 4.98 =5	55	1.67
3rd zone	$j_o r^2$ - 8.42 =8	56	1.60
4"' zone	$j_o r_3$ = 14.	56	1.75
5"' zone	$j_o r_4$ = 24.	55	1.71
6E" zone	$j_o r^5$ = 40.	49	1.71
7"' zone	$j_o r^b$ = 68.68 =69 which	69	1.68
	exactly the last existing		

Note on Statistical Regularity and Underlying Causes and models for the Bradford Phenomenon

Bradford's law is one of a number of similar patterns observed in a variety of situations. In bibliometrics, two other laws relating to author productivity of scientists and word frequencies in texts that go by the name of Lotka's Law and Zipf's law have a similar pattern where `a few *sources* produce many *items'.* Similar patterns have also been seen in economics and geography in the distribution of wealth or populations in cities. One might like to know why a ranked list of journals or citations, money, populations of cities, etc.) follow similar patterns? It is beyond the scope of this chapter to discuss these aspects. However, the present understanding is that these similarities are statistical in nature and that they arise under conditions of *cumulative advantage.* This term means that the underlying processes reinforce each other. This is sometimes also called the "success breeds success" principle. The empirical relationship observed by Bradford has also been sought to be expressed through a variety of models. Some of these features will be discussed in another lesson on Rank frequency models.

6.11 : Summary

Empirical bibliometric studies show that a high number of relevant articles in a subject field concentrate in a few journals, while the rest of the articles are scattered over a large number of journals. Scattering of articles among journals shows that for a given subject most of the journals contribute only one article each, a few journals contribute a few articles each and very few journals contribute a large number of articles. The journals can be ranked in order of their productivity. Bradford found after dividing the journals into three zones of decreasing productivity each yielding the same number of articles that the number of journals in each zone increases geometrically.

Bradford's Law serves as a general guideline to librarians in determining the number of core journals in any given field. It states that journals in a single field can'be divided into three parts, each containing the same number of articles: 1) a core of journals on the subject, relatively few in number, that produces approximately one-third of all the articles, 2) a second zone, containing the same number of articles as the first, but a greater number of journals, and 3) a third zone, containing the same number of articles as the second, but a still greater number of journals. The mathematical relationship of the number bf journals in the core to the first zone is a constant n and to the second zone the relationship is n^2. Bradford expressed this relationship as 1: n: n2.

Bradford expressed his law both as a verbal and as a graphical formulation. The graphical Bradford analysis involves the following steps: identifying the articles, listing in rank order the journals containing the articles, plotting the cumulative amounts of articles on the y-axis and the natural logarithm of the cumulative journals on the x-axis- and determining the zones from the graph.

7.0 : Objectives

After reading this Unit, you will be able to:

- explain what is a mathematical model;
- understand the ways in which a numerical data can be organized using frequency distribution;
- apply the approaches followed in analyzing data available in frequency distribution format;
- understand the concept behind a bibliometric distribution;
- indicate the characteristics of size-frequency and rank-frequency distribution; and
- carry out modeling using size-frequency and rank-frequency approaches

7.1 : Introduction

In science, one usually unfolds the nature through observations and experimentations. If certain observations are found to occur repeatedly, then it may either be just a chance or they could be some underlying regularity in this phenomenon. The aim of a scientist is to identify and analyze the pattern of such events in nature. Symbols, signs or mathematical equations can represent the overall description of this pattern or regularity. Such a description is called modeling. A model, therefore, consists of symbols or signs related through certain mathematical operations. It consists of unspecified constants known as parameters, which represent particular phenomena through a model; its symbols are assigned specific meanings. The process of representing an idea, context, concept or phenomena through a model is called modeling. In modeling the observations in the form of regularity are identified first and then expressed in some mathematical terms. During the study of a process, if the observation shows a clear pattern, then a mathematical function that can serve an empirical model of the process is suggested. By analogy, one could be benefited by another system whose behavior shows correspondence to the present case. The model essentially abstracts and reproduces certain essential characteristics and features of the phenomena to be described. What considerations and features are to be included or discarded depends upon the degree of complexity to be handled, the actual requirement and the final use to which the model is to be put.

In reality, since only essential characteristics are included, the final picture we draw from the model may be distorted. The extent of distortion may be judged in a statistical sense through the degree of fit of actual data to the theoretical model.

Testing/Application of Models

Let us study the following set of numbers

1, 5, 6, 8, 9, 12, ... (Set 1)

1, 2, 3, 4, 5, 6, 7, 8 .... (Set 2)

2, 4, 6, 8, 10, 12, ... (Set 3)

2, 4, 8, 16, 32, ... (Set 4)

You will find that:

There is no regularity in the numbers in Set 1 while there is regularity in the set of numbers represented by Set 2, Set 3, and Set 4.

The regularity in Set 2 can be represented by a symbol "n", where n represents a natural number 1, 2, 3, 4,

The regularity in Set 3 can be represented by a symbol "2n", where n represents a natural number.

The regularity in Set 4 can be represented by a symbol " 2^n ", where n represents a natural number.

If a set of numbers is given to you, you can easily find out whether that set follows a regular pattern as depicted by symbols "n" , "2n" and "2^n "in above examples.

Similar to the symbols "n", "2n", "2^n" described above, a large number of mathematical models are available in literature (described later), and one has to find out whether the set of data available with him fits any of those models.

The applicability of a model to a given data set or the fitness of a data set in a model is generally checked by the following two methods:

a) Graphical Test, and
b) Statistical Test

In practice, however, it is difficult to identify a theoretical model that describes directly completely the given set of data or in other words, the given set of data does not fully fit in any model. The process of data fitting usually consists in finding those values of the parameters of a model that give closest approximation to the data. The extent of distortion may be judged in a statistical sense through the degree of fit of the given data set to the theoretical model.

7.2 : Representations and Organization of Numerical Data

You must have organized the data in tabular form or graphical form. However, when data are too large, it can be represented by forming groups. For example, if you want to find the number of people wearing glasses in a population of Delhi, then-it would be easier to find the number of people wearing glasses above the age of 60, which could be very high, say every 9 out of ten, and the number would be very low for people below the age of 20 (say one out of 20).

7.2.1 : Frequency Distribution

A frequency distribution is the depiction of the number of times a variable occurs in a sample or population. The frequency distribution basically summarizes and compresses the data by grouping it into classes and recording how many data points fall in each class. A frequency distribution is the foundation of descriptive statistics. It should be constructed virtually for all data sets. They are especially useful whenever a broad, easily understandable description of data concentration and spread is needed

A frequency distribution has two components. One component is the scores on the concerned variable. The other component is the frequency of each score. If the data are arranged according to the size or magnitude of the variable in class intervals with a number corresponding to each showing the frequency of occurrence of the item in the class, the series formed is called a frequency distribution or simply distribution.

Example 1

Here are the sample authors and the number of papers they have contributed during one year in a specified scientific journal. Thus, author A has contributed one paper, Author B, one paper, Author C, 2 papers, author D, 10 papers, and so on:

Table 1: No. of papers contributed by different authors

1	5	1	1	1	1	8 7	7
1	1	1	2	3	2	3 4	4
2	2	4	1	1	4	2 1	1
10	1	3	1	1	3	1 1	1
3	2	1	1	1	5	5 2	2

It is hard to get a feel of this data in this format because it is unorganized. To construct a frequency distribution you should first identify the lowest and the highest values in the data set. We do it because we want to be sure that each value in the list fit in one of our categories. In the present case, the lowest value is 1 and the highest is 10.

The categories (or class) here are 1, 2, 3, 4, 5, 7, 8, and 10. For constructing a frequency distribution, categories are put into numerical order (conventionally in descending order) and the number *of* times each category occurs is counted. The data in Table 1 could thus be grouped as depicted in Table 2, which denotes the frequency distribution.

Table 2: A frequency distribution

Class Number of Papers	Frequency Number of Authors
10	1
8	1
7	1
5	3
4	3
3	5
2	7
1	19
Total	40

7.2.2 : Cumulative Frequency Distribution

The word "cumulative" generally refers to some sort of total. A cumulative frequency distribution is a way to list number of values that fit into the first category, the first two categories, the first three categories, etc. and finally the last category. From the data in Table 2, the cumulative frequency distribution can be generated as depicted in Table 3. The number of authors who published 10 papers is one and the number of authors who published 8 papers is also one. Therefore, the total number of authors who published 8 and above papers is 1+1 i.e. 2, and so on.

Table 3: Cumulative frequency distribution

Class	Frequency	Cumulative frequency
No of paper	No. of authors	Cumulative number of authors
10	1	1
8	*1*	(1+1) = 2
7	1	(2+1)=3
5	3	(3-i-3)=6
4	3	(6+3)⁼9
3	5	(9+5)=14
2	7	(14+7)=21
1	19	(21+19)=40

7.2.3 : Grouped Frequency Distribution

The above data if collected for 10 years or 50 years would become very large and will have to be compressed for a proper depiction. This compression of data is done by forming groups (called class intervals). Instead of looking at individual score, we would form groups of 1-2, 3-4, 5-6, 7-8, 9-10 papers. We could even have broader groups i.e., just two class intervals, viz.1-5 and 6-10 papers. Following points are considered while constructing class intervals:

i) A set of class intervals should be mutually exclusive. This means that categories should not overlap;
ii) All intervals should be of same width; and

iii) Intervals should be continuous through the distribution

For the data given in Table 3, the constructed grouped frequency distribution is shown in Table 4.

Table 4: Grouped frequency distribution

Class	Frequency
No. of papers	No. of authors
1-5	37
6-10	3

7.2.4 : Relative Frequency Distribution

The relative frequency distribution or percentage frequency distribution is the simple frequency distribution expressed as proportion of the total number of scores., In other words, their frequency is expressed relative to the frequency of other score. To calculate the relative frequency, the frequency of a score is divided by the frequency of all scores. So one should first calculate the total frequency and then divide each individual frequency by this total. For data in Table 3, the total frequency is 40. The relative frequency in this category I would be 1/40 (i.e. 0.025) and the percentage would be 0.025* 100 = 2.5. For the data given in Table 3, the relative frequencies are shown in Table 5.

Table 5: Relative frequency of authors

Class No. of papers	Frequency No. of authors	Relative frequency No. of authors	Percentage of authors
10	1	0.025	2.5
8	1	0.025	2.5
7	1	0.025	2.5
5	3	0.0075	7.5
4	3	0.075	7.5
3	5	0.125	12.5
2	7	0.175	17.5
1	19	0.475	47.5

7.3 : Different Approaches Used in Data Analysis through

Almost all bibliometric models relate, in a simple functional form, one variable witl another variable. The basic variables of these models are the number of journal and the number of papers, number of authors and their contributions, the words ant their frequencies, etc. Using the independent variables, we can classify then approaches into two basic types:

- Size-Frequency
- Rank - Frequency

After arranging the journals in a decreasing order of productivity, a size - frequent; distribution is obtained for the number of journals containing a fixed number o articles each. On the other hand, a rank - frequency table can be constructed for the number of articles associated with a journal of fixed rank.

The size - frequency approach gives emphasis to the journals of least productivity In mathematical terms, if g (p) is the number of journals having p papers, we speak of a size - frequency relationship. If f (r) is the number of papers in a journal o rank r, we speak of a rank- frequency relationship. In the former we can estimate the number of journals, given the number of papers; in the latter we can estimate the number of papers given the rank of the journal.

7.3.1 : Size - Frequency Approach

The size-frequency approach models the frequency of source f(x) as a function o the number of items x. It allows the use of traditional statistical techniques but ha certain disadvantages. These distributions directly represent the data for analysis and their parameters are fully explicit. They contain more information and are much simpler to use in practical context. These distributions can be analyzed and interpreted, without dependence on any statistical theory or underlying probabilistic assumption: The sources with the largest number of items correspond to the maximum value of x in the tail of the distribution. Thus, the journals with the maximum papers, the words with the highest occurrences and the authors with the maximum papers are not in main part of the distribution. If the goodness-of-fit test is performed on sue] data, the observations in the tail are placed together in one group and a change ii the magnitudes in the tail does not affect the resulting test. Even if the Kolmogorov- Smirnov test is used the tail is at values of the cumulative distribution function which are very close to one. The absolute difference therefore is very small through modeling one can predict the number of journals or authors contributing fixed number of articles.

Example 2

To understand the notions of the size-frequency approach, let us consider the following example. If *f(n)* denotes the number of journals contributing exactly t articles each to a particular subject field such that the total number of observed journals is *J = E f(n)* and the total number of observed articles is *N = n* E f(n,* (Table 6). This table relates the observation (the articles) with a class (a journal) The modeling problem is to find a mathematical equation, which relates f(n) with n.

Table 6: A frequency - size distribution of the number of journals f(n) contributing n articles each

N	*f(n)*	*N * f(n)*
No. of articles	No. of journals	Total number of articles
1	102	102
2	25	50
3	13	39
4	2	8
5	7	35
6	1	6
7	3	21
8	3	24
9	1	9
10	2	20
13	2	26
15	1	15
18	1	18
22	1	22
Sum	J=164	N=395

Source: S.C.Bradford. Sources of information on specific

Subjects Engineering 137(934): 85-86.

7.3.2 : Rank - Frequency Approach

This approach is often used in bibliometrics. It models the number of items g(r) as a function of the rank of the source r, as ranked by the number of items contributed by the source. The rank-frequency tabulation reverses the order of the size-frequency tabulation, and gives priority to the most proactive authors or journals. This type of modeling approach can be used to predict the number of articles contributed by the journal or author of a given rank.

To begin, let us understand what a rank represents in informetric studies. Looking at quantitative studies, we can assign three meanings to the word "rank."

i. Ranks are values of an ordinal variable, such as preference ("rate your professor's competence on a seven-point scale"). In this case, the invariant property is order, so that any transformation of the original values, which preserves order, is admissible.

ii. Values of a quantitative (measured) variable are replaced by their sample ranks in order to apply certain non-parametric statistical tests such as the Mann-Whitney or Kruskal-Wallis tests. The transformation is applied primarily so that the distribution of a test statistic under the null hypothesis may be determined exactly.

iii. Ranks represent cumulative numbers of sources when these. are arranged in decreasing order by number of items. This is clearly the meaning used in info-metric study,

Ties are handled in different ways, according to the type of rank. For the first type, ties are given the same value, the actual number again being arbitrary, so long as it preserves the order. For the second type, the distribution of the test statistic can be adjusted for ties. These are represented by average rank. For example, if the third and fourth smallest values are tied, they are each assigned the rank 3.5. For the third meaning of rank, ties are replaced. by. their maximal rank. For example, if the third and fourth most, productive journals have the sample number of papers, they are both assigned rank 4. The last approach is most frequently used.

To understand the rank-frequency approach, consider again the example given in Table 6. Near the bottom of the table, there is one journal-contributing maximum (twenty two). This journal is assigned the Rank 1. The next productive journal, which contributed eighteen articles, is assigned rank 2. This is continued, resulting in the generation of a rank - frequency distribution (Table 7), where g(r) is the number of articles contributed by the journal of rank r. Note that there are two journals contributing 13 articles each, and each assigned Rank 5, the "maximum-rank" assignment method which is used in case of ties. , If we assign the Rank 4 to each of these journals, then we are using a "minimum-rank method"; there are also random-rank and average-rank methods.

Table 7: A frequency - rank distribution of the number of articles g(r) contributed by a journal of rank r

Publications J	Journal $f(n)$	Total Publications $J * f(n)$	Rank R	Cumulative publications $g(r)$
22	1	22	1	22
18	1	18	2	40
15	1	15	3	55
13	2	26	5	81
10	2	20	7	101
9	1	9	8	110
8	3	24	11	134
7	3	21	14	155
6	1	6	15	161
5	7	35	22	196
4	2	8	24	204
3	13	39	37	243
2	25	50	62	293
1	102	102	164	395

Some More Examples

Some more examples of applying size-frequency approach and rank-frequency approach are given below:

Example 3

To understand the notion of size-frequency approach, consider the following example. Let *f(n)* denote the number of authors contributing *n* articles each to a particular subject field such that the total number of observed authors is *A Ef(n)* and the total number of observed articles is *N = n * E f(n)* (Table **8).** This table relates the observation (the articles) with a class (a author). The modeling problem is to find a mathematical equation relating *f(n)* with *n.*

Table 8: A frequency - size distribution of the number of journals *f(n)* contributing n articles each

n	f(n)	*n* * f(n)
Articles	Journals	Total articles
1	180 .	180
2	29	58
3	15	45
5	4	20
10	1	10
Sum	J=229	N=313

In Table 8, near the bottom of the table, there is one author, *who* contributed ten articles. This author is assigned the Rank 1. Then, there are four authors, who have contributed 5 articles each. They are each assigned Rank from 5. Then there are 15 authors, who have contributed 3 articles each. Each *is* assigned Rank 20. This is continued resulting in the generation of a frequency distribution table (Table 9), where g(r) is the number of articles contributed by authors of rank r. In the case of ties, we use the "maximal-rank" method.

Table 9: Rank - frequency distribution of number of articles g(r) contributed by the author of rank r

Publications N	Authors f(n)	Total PublicationsN*f(n)	Rank R	Cumulative publicationsg(r)
10	1	10	1	10
5	4	20	5	30
3	15	45	20	75
2	29	58	49	133
1	180	180	229	313

Example 4

To understand the notion of size-frequency approach, let us consider the followini example. Let *f(n)* denote the number of authors contributing exactly *n* articles eac to a particular subject field such that the total number of observed articles is J= ∑n* *f(n)* and the total number of observed authors N=∑n* *f(n)* (Table 10). This table relates the observation (the articles) with a class (a author). The modelling problem is to find a mathematical equation relations *f(n)* with *n.*

Table 10: A frequency-size distribution of the numberof authors f(n) contributing n articles each

n.	f(n)	n.* f(n)
No. of articles	No. of authors	Total No. of articles
1	85	85
2	37	74
3	17	51
4	15	60
5	3	15
6	3	18
7	1	7
8	2	16
10	1	10
11	1	11
12	1	12
15	1	15
21	1	21

In Table 10, near the bottom of the table, there is one author who contributed 21 articles. It is assigned the Rank 1. The next most productive author who contributed 15 articles is assigned Rank 2. This is continued, resulting in the generation of a frequency-rank distribution table (Table 11), where g(r) is the number of articles contributed by the author of rank r. Note that there are two authors who contributed 8 articles each. Here each is assigned Rank 7, using the "maximum rank" method, which is used in case of ties here. Similarly there are 3 authors, who contributed 6 articles each, and each is assigned Rank 11; three authors who contributed 5 articles each are assigned Rank of 14; fifteen authors who contributed 4 articles each, are assigned Rank 29 each, and so on.

Table 11: Rank Frequency distribution of number of articles g(r) contributed by the author of rank r.

Publications N	Authors f(r)	Total PublicationsN*f(r)	Rank R	Cumulative publications g(r)
21	1	21	1	21
15	1	15	2	36
12	1	12	3	48
11	1	11	4	59
10	1	10	5	69
8	2	16	7	85
7	1	7	8	92
6	3	18	11	110
5	3	15	14	125
4	15	60	29	185
3	17	51	46	236
2	37	74	83	310
1	85	85	168	395

7.5 : Summary

An attempt has been made in this Unit to explain the concept behind a mathematical model, ways in which data generated from library and information activities can be organized. The various approaches used in data analysis are also described. Various features of size-frequency and rank-frequency approaches for analyzing data have been explained. A selected list of various bibliometric models *using* these two approaches are indicated.

Unit-8:Informetrics Phenomena

8.0 : Objectives

After reading this Unit you would be able to:

- understand the Informetrics Phenomena and its specialized terminology;
- appreciate the usefulness of informetric laws and their applications in practice as well as in research in the field of library and information science; and
- recognize the occurrence of the phenomena in other subject fields as well.

8.1 : Introduction

Informetrics, a generic term encompassing librametry, bibliometrics anchseientometric is the "use and development of a variety of measures to study and analyze several properties of information in general and documents in particular. This Unit highlights the development of the applications of quantitative methods in library and information studies. A knowledge of informetrics phenomena and application of the related lam can help in selecting and building information source collections, improving library and information services to meet the needs of users, and more efficient management of library and information systems and services. In other words, the knowledge enables fulfilling the implications of S. R. Ranganathan's Five Laws of Library Science more effectively, for instance, information use, convenience and economy of effort of the user, and efficiency and effectiveness of system management, all this being achieved even as the information and user environment constantly change.

In general, informetrics seeks to develop statistical and mathematical techniques evaluate and improve the efficiency of information services and their utilization. In the area of collection development in library and information centres, informetrics applications help in the selection of core periodicals for subscription, selection of infrequently used books and periodicals that may be withdrawn and placed in depositories to meet occasional demands, archiving of less used electronic resource etc. Informetric techniques can also be used in evaluating how well a library) / information service is performing, say in comparison with other centres; or how frequently and extensively the output of scientific papers of an author, an institution or a country is cited by researchers in the field within and outside the institution / country. Of course such measures need to be used in conjunction with other qualitative measures.

"This resurgence of interest in informetric studies] comes because of the growing demands on library and cognate services. Far more people are now working at, and paying for, library and documentary work than earlier. Moreover, workers are now quite as numerate as they are literate, sometimes more so.... Also we can now delegate heavy computation to automatic machines. Above all, this increased interest has been made possible by those who have collected and presented essential data about who and how many and in what way and when and where and by what name." (Fairthorne, 1969).

8.4 : Informetrics Phenomena in Science

Application of statistical methods to organization of science and productivity analysis thereof is usually termed Scientometrics. Some researchers consider it as part of informetrics as several of the bibliometrie and related phenomena mentioned in the preceding sections can also be observed in the field of science. One of the classical laws in scientometrics is Lotka's law mentioned above. Some of the other laws are:

1) Distribution of significant contact between scientists in a particular subject field and those specializing in other fields, can be expressed as a Poisson Distribution:

$$f(x) = \frac{e^{-\lambda}\lambda x}{x!} \; x = 0,1,2,3,...$$

$$x > 0$$

where **x** is the number of significant contacts between scientists in a field and "outsiders"; l is the mean number of significant contacts over the population of scientists; **f(x)** is the probability of a scientists having x significant contacts with outsiders.

1) Distribution of citations to a paper during time t_o and t follows a Poisson Distribution.

2) De Solla Price conjectured that "half the successes (of scientific productivity) were due to the highest scoring elite comprising √P of the individuals, that is √P authors will produce at least half the total papers published by the population **P**·

Goffman's Epidemic Theory [1966] shows that the process of diffusion of information (e.g. innovations) is analogous to the process of spread of epidemics and, therefore, the diffusion process can be represented by ferential equations

$$\frac{ds}{dt} = \beta SI - \delta S + \mu$$

$$\frac{ds}{dt} = \beta SI - \lambda I + \upsilon$$

$$\frac{ds}{dt} = \delta SI + \gamma I$$

Where **S, I,** and **R** are continuous functions of the real variable **t;** β is the rate of infection (acceptance) and λ (γ) is the rate at which susceptible and removal, and g(v) is the rate at which a new supply of susceptible enters the population.

Neelameghan and others [1970] have shown that the pattern of duplication (once, twice, thrice, etc.) of the discovery of antibiotics follows a modified Poisson Distribution. This practical finding was the first confirmation of a conjecture by De Solla Price. The Square-Cube Law *from* Biology (also applied in Management) is also applicable in this case and provides a better prediction of the duplication [Neelameghan, 1971].

8.5 : Practical Applications of Informetrics

In general, as already mentioned, informetrics seeks to develop statistical and mathematical techniques

to evaluate and improve the efficiency of information services and their utilization. In the area of collection development in library and information centres, informetrics applications help in the selection of core periodicals for subscription, infrequently used books and periodicals that may be withdrawn and placed in depositories to meet occasional demands, archiving of less used electronic resources, etc. Informetric techniques can also be used in evaluating how well a library / information service is performing, say in comparison with other centres in the country; or how well the scientific output of an institution or a country is cited by researchers in the field within and outside the institution / country. Of course such measures need to be used in conjunction with other qualitative measures.

Informetric studies have mainly been used to evaluate the extent to which different types of users and their needs are satisfied. In recent years reports of more analytical rather than descriptive studies are being published. These studies can be categorized as follows:

1) To measure the adequacy of a library's collection for the present and expected future programmes of an institution. A number of such studies were reported in the 1950s and in the 1960s.
2) To discover mathematical and statistical models for various phenomena experienced in library and information work.

Hopefully informetric studies will help to improve the efficiency and effectiveness . of library and information services, even as the information environment changes,

as envisaged in Ranganathan's Five Laws of Library Science.

8.6 : Summary

This Unit presented a brief overview of the concept of Informetrics Phenomena. Discussed the usefulness of understanding and applications of statistical methods in library and information work. Following a note on the historical development and terminology of informetrics, the principal classical bibliometric / informetric laws are considered, covering studies in literature growth patterns, obsolescence rate patterns, success-breeds-success phenomena and their applications in science. This is followed by a discussion of the practical applications of informetrics in library and information work and service.

The Bradford and Zipf distributions discussed above are, a class of empirical hyperbolic distributions. A characteristic of hyperbolic behaviour is that the product of fixed powers of the variables is constant. Such behaviour has been noticed in different fields - economics, meteorology, psychology, linguistics etc. And explanations appropriate to the particular fields have been provided. As a result, different names have been given by the researchers in the different fields. Fairthome (1969) attempted unification of the statistical distribution patterns in the framework of the researches by Mandelbrot (cited in Fairthorne (1969).

Unit-9:Analysis of Library Related Data

9.0 : Objectives

After reading this unit you would be able to:

- get an insight into how to analyze library data, and their better utilization for the benefit of the library users;
- perform citation analysis and ranking of scientific periodicals to the best advantage of its users;
- understand the scope of librametrics or librametry, and how to quantitatively analyse various library activities;
- understand the genesis, definition, scope, purpose and application of bibliometrics, and how bibliometric analysis helps library and information personnel to provide better library services to its users; and
- learn the definition, scope, and, usage of informetrics; and scientometrics.

9.1 : Introduction

The word **"metrics",** when suffixed to a field of activities, it indicates application of mathematical / statistical measuring techniques to that field, generating a new sub-field out of fusion of the terms. The word **"metrics",** derived either from the Latin or Greek word, *"metricus ",* or *"metrikos"* respectively means measurement. For example, when the term "metrics" is suffixed to terms like biology, sociology, psychology, economics, etc., the resultant fusion of the pair of concepts has generated more new areas, such as biometrics, socio-metrics, psychometrics, and econometrics, etc., where various standard techniques of mathematics and statistics have been extensively used for better understanding of these subjects and their applications.

Likewise, the terms librametrics, bibliometrics, informetrics, and scientometrics have also been derived out of the fusion of the term **"metrics"** with library, bibliography, information, and science respectively. Librametrics, bibliometrics, informetrics, and scientometrics, are well-known measuring techniques in library and information science. The concepts connoted by these terms overlap with each other, and therefore need elucidation for a clear understanding of their manifold applications in library and information science.

9.2 : Necessity for Analytical Studies in Libraries

Libraries may be broadly classified into two types:

- **General Libraries** (including academic, and public libraries), and
- **Special Libraries**

This categorization is made on the basis of the users and their information needs. Karlyle makes a clear-cut distinction between these two types of libraries. In a general library, the users being generalist readers, the librarian has to spend a major portion of his resources to purchase macro-documents like textbooks, fiction, popular weeklies, etc. On the other hand, in a special library, the users being specialist readers, the librarian spends a significant amount of his resources in procuring micro-documents in the form of periodical publications, etc. In a general library, the librarian stresses most on functions like purchase, accessioning, classification, cataloguing, circulation, binding, and so on. On the other hand, in a special library, in addition to the types of work just mentioned, the librarian acts as a disseminator of knowledge by providing various types of services, such as, Current Awareness Service (CAS), Selective Dissemination of Information (SDI) service, Abstracting and Indexing Services (AIS), Bibliographic Service (Literature search service), Document Delivery Service, Translation Service, Reprographic Service and so on. By organizing these services a special library caters to need-based information to its users. Rabindranath Tagore's view of a library is that the bigness of a library is measured not by its size, but by its use.

Today the library is no more viewed simply as a store house of books. The output of scientific literature started increasing, especially in the second quarter of the 20^{th} century. New journals in specialized branches of science keep appearing, and established periodicals either have increased the number of its volumes per year, or split into different titles covering narrower subject areas in order to cope with the flow of ideas and the pressure to disseminate them as widely and as quickly as possible.

From the post-second World War period, scientists all over the world are facing a 52 *Literature* or *Document Explosion. As* a result scientists in general face the Herculean task of keeping themselves abreast of the current developments in their areas of research. This has important impact on the library. As a first measure, many of the closed access libraries have been converted into open access ones, where the library users can go straight to the library racks where all the library materials are arranged in a preferred logical sequence, and select reading materials. Librarians have become disseminators of information. They have acquired expertise to collect, collate, organize, store, retrieve, and disseminate information `pin pointedly, expeditiously, and exhaustively' to the users, either on demand or in anticipation, with a view to `Save the Time of the Reader'. Side by side, librarians have introduced to conduct regular analytical surveys of the activities of different sections of the library, with a view to streamline various library services. `Library is a Growing Organism', this concept of Dr. S. R. Ranganathan calls librarians to adopt a dynamic weeding policy, and regulation of the day-to-day inflow of documents in the library.

In order to develop and maintain an active and useful collection of information resources the library needs to analyze users' interests and requirements on the one hand and the information materials acquired to satisfy those needs, on the other. Such regular analytical studies help the librarian to identify the frequently used, less frequently used and unused documents. By withdrawing the unused documents permanently from the library racks, and transferring infrequently used documents into the stack room, it is possible to save on the shelf space. It will also help decisions regarding discontinuing subscriptions of less used periodicals, and thereby saving money for acquiring new and more useful periodicals. Since library budgets do not generally increase in parallel with the increase in the volume and types of information, the library has to adopt a judicious and dynamic procurement policy. In a research library, predominant use of scientific periodicals, calls for a well planned journals subscription policy, so that library budget is spent for the best advantage of its users, and to ensure that the researcher is spared the labour of wading through publications of relatively little value. To ensure this, the librarian evolves his subscription policy by consulting ranking lists of scientific periodicals. These ranking lists are prepared through the process of citation counting. To make this ranking list effective and accurate, the lists should be revised at regular intervals, as with the efflux of time, and due to shift of research interest, a highly ranked journal may take a lower rank, while a lowly placed journal may occupy a higher position. Similar analysis should be undertaken in respect of readers' statistics/users' survey to identify information needs of the users. These findings together with the list prepared from the analysis of used documents, should be the actual guide for journal selection and subscription.

9.3 : Citation Counting: A Versatile Tool for Journal Se

A common tool for the selection of serial publications for a research library is the list of most cited serials. The term *"serials"* here denotes publications brought out in different parts intended to be published for an indeterminate period of time, generally at regular intervals, not more than once a year, though irregularity in intervals is not ruled out. The possibility of grading scientific serials was first conceived by P.L.K. Gross and E.M Gross in 1927. They applied a simple statistical investigation to the grading of scientific serials according to their relative importance, based on citation counting from some primary source journals selected for the purpose. Similar surveys were made by E.S. Allan (1929) of periodicals in mathematics, J.K.McNeely and C.D.Crosno (1930) of periodicals in electrical engineering;, and P.L.K. Gross and A.C. Woodford (1931) of periodicals in Geology, Subsequent studies include those in the field of Medicine by R.Jenkins (1931), and by J.Gregory (1937), in Biochemistry by H.N.Henkle (1938), and in Physiology by E. Broadman (1944). In 1956, C.H.Brown presented, in the Monograph No. 16 of the Association of College and Reference Libraries, Chicago, lists of most cited publications in Mathematics, Physics, Chemistry, Geology, Physiology,' Botany, Zoology, and Entomology. From India in 1970, Sengupta presented two comprehensive ranked lists in for Medical Sciences and Biomedical Sciences respectively, from Indian scientists' point of view.

All these ranking lists of scientific periodicals in different disciplines were prepared applying the Gross methodology based on counting citations from a set of source journals selected on the basis of their coverage, excellence, reputation and circulation. All citations found at the end of articles, within the text of the articles and in footnotes are counted.

J.Biol.Chem		*Am. J. Med*
YEAR	CITATIONS	TOTAL CITATIONS
1998	卌 卌 卌 卌 卌	25
1999	卌 卌 卌 111	18
2000	卌 卌 卌 卌 卌 卌 111	33
2001	卌 卌 卌 卌 卌 111	28
2002	卌 卌 1111	14
	GRAND TOTAL	118

Fig. 1: Specimen Tally Card

Proceedings of international congresses, colloquia, meetings, etc., are considered as serials, provided a permanent organization exists for arranging their publication, otherwise they are classed as non-serials. Unpublished data, personal communications, theses, papers in press or awaiting publication are noted separately and treated as non-serial citations. All the names of cited journals are noted in 5" x 3" tally card or slip, and arranged in alphabetical sequence. A specimen of the card with all entries is shown in Fig. l.

The title of one of the source journals, *Journal of Biological Chemistry*. is noted at the top of the left column, wherefrom the citations of the cited serials are collected. At the top of the right hand column the title of one of the cited serials *(American Journal of Medicine)* is noted. The *modus operandi* of citation counting, though laborious, is relatively simple. All references appended at the end of each article of the source journal are year-wise noted in separate columns *(vide* specimen card). The number of citations collected from all the source journals, for each individual cited journal is then added up and the overall ranking list is made on the basis of the total number of citations received by each of the journals cited in the source journals. The cited journal, which receives highest number of citations, occupies the first rank. The journal, which receives second highest number of citations, is ranked second and so on. Table I presents a specimen ranked list.

RANK	NAME OF JOURNAL	TOTAL NUMBER OF CITATIONS
1	*J.biol.Chem.*, U.S.A	2032
2	*J.molec.Biol.*, U.K.	1985
3	*Nature, Lond.*, U.K.	1975
4	*J.Cell.Biol.*, U.S.A	1786
5	*Proc. natn. Acad. Sci.*, U.S.A	1680
	. . .	. . .
660	*Danish med. Bull.*, Denmark	07
	GRAND TOTAL	9465

Table 1: Ranked List of Periodicals in the Field of Biochemistry

In 1956 Brown suggested that ranking of scientific periodicals in a particular discipline can also be made by selecting appropriate secondary periodicals as source. He foresaw that for the frequency of abstracts of a periodical reflects the frequency of its citations in primary journal, and is, therefore, an index of its relative importance at the time of citation. Although the method of citation counting was developed with a view to getting a factual approach to the selective value of scientific periodicals in the discipline concerned, the method has limitations. Ranking lists of journals thus prepared cannot be taken as the final word for their selection. In case of doubt, before finally deciding to purchase, views of experts may be sought. Further, this process of evaluation should be a continuous one, as the rank of individual journals is bound

to change with the efflux of time. Due to shift of research interests, low-placed journals may occupy higher ranks, while a highly ranked journal may come down to a lower position. It is, therefore, desirable that such lists should be prepared, checked and compared at regular intervals and the selection policy reviewed in the light of the latest available lists.

9.4 : An Alternative Method of Citation Analysis

Citation analysis devised by Gross and Gross in 1927 was considered as the on authentic method to rank scientific periodicals up to the later part of 1960s. However that Gross's method has some in-built pitfalls, such as, bias due to language and country of origin of the source journal. Moreover, Gross's method is not only very, laborious but also erroneous, if the source journals are not properly chosen. The limitation of Gross's method of citations counting from primary source journals arise from the difficulties in the selection of a sufficiently representative group of source journals in order to obviate the above mentioned bias. The country of origin source journal, and the language used in its text have unavoidable bias on the journals cited. The selection of suitable source journals has always been a problem to the compiler of such lists. The only way to eliminate such bias is to increase the number of source journals so as to ensure a wide geographical and linguist coverage. The labour involved in citation counting, militates against selecting too large a. number of source journal; at the present time, the number of primary journals in any field of scientific knowledge, which in consideration of their standard and coverage cannot be omitted from any representative list of source journals, h grown so large that the method of counting citations from primary source journal is becoming time consuming, tedious, and impractical. Further, the grown specialization within different scientific disciplines, the opinion of experts, whom tl compiler considers most competent for selection of source journals, is tending become increasingly divergent even within a country. It is because of these limitation Gross's method of ranking scientific periodicals, generates false signals in the ranking table, and thus, cannot be taken as an authentic tool for ranking scientific publication (Sengupta, 1972)

The second method suggested by Brown during 1950s based on counting tl frequency of abstracts from the concerned secondary periodicals as an alternate' to Gross's method, is also subject to certain pitfalls. The declared objective reputed secondary periodicals like, *Chemical Abstracts, Biological Abstracts, Ind Medicus,* etc. is to ensure comprehensiveness in their respective areas of interest; and abstracting has ceased to be selective. In consequence, the results of counting citations in the secondary periodicals will be determined only by the number P articles published in a journal, during the period of time covered.

9.4.1 : Selection of New Source Journals to Eliminate Bias

Considering the difficulties and limitations of Gross's and Brown's methods, Sengupta attempted to develop a simpler and more objective method of evaluation by citation counting based on the impact of individual journals in the progress of knowledge in a particular field. Instead of Gross's method of counting references from a set of primary journals counted citations from the bibliography at the end of each article of Annual Reviews (Sengupta, 1972), The Annual Reviews on different scientific disciplines are published by Annual Reviews, Inc, Palo Alto, California, and have established themselves as reputed primary serial publications with an international coverage. The reviews appearing in them selectively focus on recent advances in their respective subject areas. Besides, the reviews are contributed by specialists drawn from different countries, so that any bias on account of nationality and language will be automatically corrected in the totality of citations. Analysis of literature in a specific discipline revealing recent advances describing, evaluating, and placing in context the more significant contributions by a combination of contributions from different experts. The resultant bibliographic completeness, and a high degree of objectivity in evaluation, support the grading of scientific serials based on the use of the Annual Reviews for source journals. From the point of view of objectivity ranking should ideally be done on the basis of impact of published work on the advancement of knowledge in a particular field. For this purpose, the Annual Reviews provide a reliable practical means for evaluating the impact of articles on scientific progress, since citations in Annual Reviews are based on relevance of a paper to the advancement of knowledge in the particular area of knowledge, and not to its relevance in the context of isolated investigation by an individual or a group, In the later case, there is a preponderant emphasis on citation describing the methodology; and primary citations relating to methodology tend to be repeated several times, which may lead to bias for a primary journal in which these papers on methodology are published.

Further, counting citations from Annual Reviews is much less laborious, and convenient, because in any particular field of knowledge covered by the Annual Reviews, counting may be confined to one or more volumes of a single source journal.

9.4.2 : Weightage Formula to Correct Citation for Post-War

Ranking of scientific periodicals, in general, is based on the total number of citations received by each cited journals, derived from a set of source journals for a specific period of time, say, three to five years. In such lists, equal weightage is given to all citations irrespective of chronology. Thus in all these ranking lists, the positions occupied by newly established periodicals may not be a true index of their demand, or usefulness to research workers in the concerned field. This follows from two considerations:

- The quantum of published literature in the new periodicals is limited by the period for which they have been in existence, which in many cases, is less than 20 years, the maximum interval for which the ranking list may remain useful. If the trend, adjudged by citations to the available literature, is towards increasing use of these journals, a correction needs to be made for the smaller quantum of literature available in them.

- Many of the post-war journals appearing in all recent ranking lists have started publication even after 1960s, and the enumerated citations of them cover in fact a variable period of less than 20 years, unlike in the case of pre-war journals.

To correct these factors (Sengupta, 1984) introduced an off-setting weightage formula, as follows:

$$C_1 = C \times [(y_2-y_1)/n \times (y_2-y_3)/n_1] \qquad (1)$$

Where, C_1 is the number of corrected citations;

C is the number of enumerated citations;

y_1 is 1946, the beginning of the post-war period;

y_2 is the year upto which the source journals are consulted for counting citations;

y_3 is the beginning year of the source journals from which counting of citations started;

n is the number of years during the period (y_2-y_1) for which the post-war journal has been in existence;

n_1 is the number of years during the period (y_2-y_3) for which the post-war journal has been in existence.

To apply the weightage formula, let us cite an example. Let *Annual Review of Biochemistry* be the source journal for the year 1968 to 1970, that is, for 3 years. Here 1970 is the year up to which the source journal *A. Rev. Biochem.* was consulted. Therefore, in the weightage formula,

$Y_2 - y_1 = 1970 - 1946 = 25$ years, and

$Y2 - {}^{y}3 = 1970 - 1968 = 3$ years.

Therefore, equation (1) becomes,

$$C_1 = C\ [25/n] \times [3/n_1] \qquad (2)$$

In this case, the factor 25/n corrects for the smaller quantum of literature available from journals starting publications after 1946, and the factor $3/n_i$ for the shorter period covered by enumeration in respect of journals published after 1968. The weightage thus given may not be considered exaggerated, because, in fact, we have still omitted the accumulation of pre-1946 literature in the case of long-established journals, like *J. Biol. Chem.*, ranked first in the original ranking list of biochemistry made by Sengupta (1973b). He preferred to neglect this factor in view of the following facts:

- the quantum of pre-1946 citations was negligible;
- that this might be off-set by the subjective tendencies on the part of authors to give rather greater attention to newly established periodicals, because of their novelty.

In any case, it is believed that the correction applied should yield too low, rather than too high, a value.

Let us apply formula (2) to a post-war journal, say, *J. Molec. Biol.* which started publication in 1959. It received 1004 citations, and ranked 4 in Sengupta's ranking list of Biochemistry (Sengupta, 1973a). Here, enumerated citations, C=1004, n=1970-1959=12, 1970 is the year up to which the source journal *Annual Rev. Biochem.* was consulted; 1959 being the year of inception of *J. molec. Biol.*, n_l=1970-1968⁼3 years, for which the source journal, *A. Rev. Biochem.* was consulted. Thus, from equation (2) we get,

$C\sim = 1004 \times (25/12) \times (3/3) = (251 \times 25)/3 = 2091.6 = 2092$

By applying Sengupta's weightage formula, the post-war journal *J. Molec. Biol.* gained citations from 1004 to 2092, ie. it has gained 1088 more citations.

9.4.3 : Three New Bibliometric Parameters to Re-Rank Scien

Limitations of the method of Gross and that of Brown have been elaborately discussed in the earlier section, and remedial measures have been suggested by Sengupta (1989) to eliminate bias due to country and language, etc. He further devised a weightage formula which corrects the small quantum of citations received by post-war periodicals, while preparing a ranking list of periodicals. Apart from all these factors, there are some other considerations, which may influence ranking lists adversely. Number of citations in relation to size of journals, and average length of papers published in journals, are two such facets, which should get due consideration while preparing a ranking list. Sengupta, (1973a; *Scientometrics,1986)* introduced the following three bibliometric parameters to counter these adverse effects:

- Scientific interest of a journal in relation to total number of articles published (D/ A);
- Compactness of information content in a scientific periodical (C/A); and,
- The scientific value of the papers in relation to compactness of presentation

$$[(D/A) / (C/A)] = D/C$$

In these three parameters, namely, D/A, C/A, and D/C:

A denotes for total number of articles published in a journal during a particular year,

C denotes total number of words published in that journal during that particular year, and

D denotes the total number of citations noted in favour of that journal from the source journal during that particular year.

Let B be the total number of pages published by that journal, say, *J. biol. Chem.,* during the year, say, 1989. The actual values of A and B are calculated by noting the number of articles, and pages published in that journal, say, in 1989. To calculate C for that journal one has to randomly select 100 pages from the source journal, say, *J. Biol. Chem.,* and counted the number of words in these 100 pages. If the total number of words in 100 pages of a journal be C', then the average number of words published per page in that journal will be C'/100. Thus the total number of words published by that journal during the year, say, 1989, is $C = B \times C'/100$, where B is the total number of pages published by that journal during the year 1989. Obviously the numerical value of C will differ from journal to journal, because of their variation in bulk, and format. The numerical value of D is known from the year-wise citation data collected from the source documents. Since all the values are known, the numerical values of parameters, D/A, C/A, and D/C can be easily calculated. When these parameters are applied to a ranking list, to re-rank them, it is expected that the re-ranked list of journals will be the best tool for selection and subscription of journals.

9.4.4 : Garfield's Method for Cito-Analytical Studies

Eugene Garfield launched Institute of Scientific Information (ISI), Philadelphia. Since its establishment, it became an active focal point for information handling, information storage, and information transfer. The year 1963 may be considered as a landmark in the history of citation analysis, because in that year, Garfield, first introduced *Science Citation Index (SCI),* which has provided a new dimension to citation analysis. SCI covers more than 3500 periodicals considered most significant journals all over the world. SCI comprises several parts, namely, Citation Index Source Index,. Permuterm Subject Index, and Corporate Index. It is a bi-monthly publication, cumulated yearly and quinquennially (that is five yearly). Quinquennial cumulation covering **the** period 1965 - 1969, 1970 - 1974, 1975 - 1979, 1980 - 1984, 1985 - 1989, 1990-1994, and decennial cumulation covering the period 1945 **-** 54, 1955 - 64, 1965 - 74, 1975 - 1984, 1985 - 1994 etc.; are also available. Half yearly cumulation, and annual cumulation in CD-ROM format are also available. Institute of Scientific Information, like SCI, also brought out *Social Science Citation Index* (SSCI), and *Arts and Humanities Citation Index* (A&HCI) in 1973, and 1978 respectively. In addition, ISI launched *Journal Citation Reports* (JCR), 1975. All these publications of ISI are considered as versatile tools for preparing citoanalytical products. By cito-analytical products, we mean those products, generally prepared by analysis of citations. Citations are generally references given by an author at the end of his/her article. The results of citation analysis are generally based on cited authors, cited journals, year of publication, etc. These results are prepared in graphical and tabular forms, and are better known as cito-analytical products. These publications provide important and useful information about the world's top-ranking scientific and technical periodicals. From these publications, especially from JCR, one can get ready-made ranked lists of scientific/technical periodicals, on the basis of total citations. It also furnishes ranked lists of journals based on Impact Factor and Immediacy Factor. It is also a storehouse of many other important information useful not only to research workers, research organizations, universities, etc. but also to librarians, who, based on ISI publications, specially on JCR, can finalize their subscription policy to the best advantage of their users. It may be noted that *Citation Index* lists citations of first author arranged in alphabetical sequence. These citations are collected from the references, generally appended at the end of a scientific/technical article, within the texts of the articles, and also in footnotes, and are arranged under the surname of the first author.

The following are the more important cito-analytical products, developed by the ISI, Philadelphia:

- ranked lists of journals by citations,
- ranked lists ofjournals by Impact Factor,
- ranked lists of journals by Immediacy Factor,
- ranked lists of journals by Cited Half-Life,
- ranked lists of journals by Citing Half Life,
- ranked lists of authors,
- citation profile of a scientist
- citation profile of a research institution, university, etc.

All these cito-analytical products have been elaborately discussed, including the limitations of these cito-analytical methods in Unit 11, Block 3, of MLIS-02, it is redundant to discuss them here again.

9.5 : Analytical Measuring Techniques for Library Data

In earlier sections, we have discussed several analytical methods useful for analysis of library related data. By applying these methods, the librarian of a special library is in a position to reorganize various library services efficiently for the benefit o the users. As for example, analytical study of library documents left by users or the tables of the reading hall of the library helps the librarian to discard unused documents from the library racks. Likewise, less frequently used documents mad be removed from the library racks to stack room of the library. This analysis o: library related data helps the librarian not only to save considerable shelf-space o: the library, but also to save unproductive expenditure by deleting the titles of unused and less frequently used documents from the subscription list. Further, proper analysis: of readers' statistics/survey, enables the librarian to know the information needs o; the clientele, which in turn helps in formulating effective procurement policy for the library. Librarians maintain regularly profile data for every individual user of the library on the basis of the feedback received from users. These profiles are analyzed to organize purposeful and personalized library services, such as Current Awareness Service (CAS), Selective Dissemination of Information (SDI) service, etc. Likewise, Inter-library Loan Service, Translation service, Abstracting service, Document Delivery, etc. may be properly planned based on analysis of users' requests.

The four measuring techniques to be discussed in the following sections are:

- Librametrics
- Bibliometrics

- Informetrics, and
- Scientometrics

9.5.1 : Librametric Analysis

The term **Librametries** has two roots, namely, Libra, and metrics. The term Libra connotes library, and metrics means measurement. The term *Librametry* was first proposed by S.R. Ranganathan. It was formally introduced by him in 1948 at the Aslib's Annual Conference, held in Leamington Spa, when he was asked by the President of Aslib. Henry Lemeisterto comment on the lecture of J. D. Bernal. Being a mathematician, Ranganathan knew that statistical / mathematical analysis is the key technology for all forecasting studies. He foresaw that library work and services involve "large numbers". He, therefore, applied mathematical and statistical techniques to streamline the day-to-day library activities for better library management, to organize need-based library services for users. Thus, as far back as 1925, as librarian of the Madras University, Ranganathan applied Librametric techniques to solve day-to-day library problems, and to provide greater flexibility in library activities and services. In 1969, at the annual seminar of the Documentation Research and Training Center (DRTC), Ranganathan (1969) in a paper described **how he** applied Librametry in the Madras University. Ranganathan, however, did not provide an exact definition of the term Librarnetry. Ravichandra Rao (1985) introduced a common definition for Biblornetrics, and Librametrics as follows:

> "Information processes and information handling in libraries, and information centers by quantitatively analysing the characteristics, and behaviour of documents, library staff, and library users."

Sengupta (1985) defined librametrics as:

> "Quantitative analysis of various facets of library activities and library documents by application of mathematical and statistical calculus with a view to seek solution to library problems".

From this definition, it is evident that the scope of librametrics not only rests on quantitative analysis of various library activities, but also it is concerned with the incoming and outgoing macro- and micro-documents of a library. Library activities and services are quantifiable by application of statistics, as many of them involve large numbers. As for example: library use statistics, statistics of used documents, statistics of incoming and outgoing documents, calculation of shelf-space of the library, annual cost of purchased materials, various library indicators, say, library performance, etc. circulation statistics, interlibrary loan statistics, statistics of foot-notes, and references, statistics of man-hour needed for each item of library work for developing a standard staff formula, statistics of deployed man-hours in different sections of the library, during different library hours, distribution of different library documents, according to different authors, and subjects, different structural aspects of the library, etc. involve large numbers, and are, therefore, quantifiable. All these suggest that application of statistics can be very effective for analyzing different aspects of a library system. It may, therefore, be concluded that the scope of Librametric studies, lies in the quantification of the characteristics and behaviour of all types of library documents, library users, library personnel, and other structural components of the library. Or in other words, quantitative analysis of the entire library system including its management and activities fall within the scope of librametrics.

Ranganthan (1969) and Neelameghan (1969) pointed that the scope of Librametric studies in a library system is manifold. Some of the areas where Librametric techniques can be profitably applied are enumerated below:

- to determine optimal number of library staff in different sections, and their judicious deployment,
- to deploy library staff to dispose of reference queries during different library hours,
- to systematize circulation of library documents,
- to develop and organize local, state, regional, and national library system,
- to determine the optimal size of a service library,

- to design library building along with its fittings and furniture,
- to distinguish between service and dormitory libraries,
- to analyze readers' queries,
- to streamline acquisition system by analysing library documents, books, monographs, periodicals, etc.,
- to determine the most helpful sequence for classification of macro- and micro-documents,
- to test the accuracy of catalogue entries by sampling techniques, and,
- to ensure proper selection of library documents.

From all these, we may conclude that basic aims and objectives of Librametric studies is to help to improve service to library users by organizing need-based library system, and its utility rests on the fact that it can provide some potential tools to ensure quantitative measurement of cost-effectiveness, and cost-benefit of investments in library systems.

9.5.2 : Bibliometric Analysis

The word *Bibliometrics* has two roots: (i) .biblio, and (ii) metrics. The term *biblio* is derived from the combination of a Latin and Greek word *biblion* equivalent to Byb1(os) meaning book, paper, which, in turn, was derived from the word Byblos, a city of Phoenicia, noted for export trade in paper. The word *metrics* on the other hand, indicates the science of meter, ie. measurement and is derived either from the Latin or Greek word, *matricus* or *metrikos* respectively, each meaning measurement.

Bibliometrics is a relatively new subject and a branch of information science. Its scope, according to O'Connor and Voos (1981), includes the study of the relationship among documents e.g. citation studies, or describing a literature. Typically, these descriptions focus on consistent patterns involving authors, monographs, journals, or subject/language". Its techniques have extreme applications, equally in Sociological Studies of Science, Information Management, Librarianship, History of Science, including Science Policy, Study of Science, and Scientists, and also different branches of Social Sciences, and Social Scientists, The techniques of bibliometrics yary from simple to complex in nature, and are not always free from controversy. The basic units of bibliometrics are all facets of written communications, such as, primary and secondary periodicals, articles, and abstracts published in them, bibliographies of articles, books, monographs, and other media of communications. Day-by-day, it is attaining sophistication, and complexity having national, international, and inter-disciplinary character. It has established itself as a viable and distinctive technique for evaluation of science based on bibliographical and citation data. Its backbone lies in its sound theoretical foundation, most effectively laid by some pioneers, such as, Lotka, Gross, Bradford, Zipf, Derek J de Solla Price, Bookstein, Mandelbrot, Brookes, Narin, Garfield, Vickery, Moravcsik, Cole brothers, Pritchard, Dobrov, Vlachy, Hulme, Fairthorne, and others. Its techniques are useful in analyzing and finding solutions to various complicated problems faced by information scientists to quantify the process of written communication.

The term *Bibliometrics is* of recent origin. It is analogous to Ranganathan's *Librametrics,* The Russian concept *Scientometrics,* FID's *Informetrics* and similar to other well established subjects, such as, *Econometrics, Sociometrics,* and *Biometrics,* where mathematical, and statistical calculus are systematically applied to study and solve problems in the related field - Library Science, History of Science, Information Science, Economics, Sociology, and Biology respectively. The term *Bibliometrics* was first proposed by Pritchard in 1969 **in** preference to the existing terminology *statistical bibliography,* as he felt that the latter term has some built-in ambiguity, and there is a fair chance to misinterpret it as bibliography of statistics. Pritchard defined bibliometrics as "The application of mathematical methods to books, and other media of communications", its purpose being "To shed light on the process of written communication, and of the nature and course of development of a descriptive (In so far as this is displayed through written communication), means of counting and analysing the various facets of written communication". According to Fairthorne (1969), bibliometrics is the "quantitative treatment of properties of recorded discourse, and behaviour appertaining it". Roland Stevens defined Bibliometrics as a "Quantitative Science", and divided it into two basic categories: Descriptive bibliometrics (productivity count), and Evaluative bibliometrics (literature usage count). Hawkins (1977) interpreted bibliometrics as "Quantitative analysis of the bibliographic feature of a body of literature". Potter (1981) defined Bibliometrics as "the study and measurement of the publication patterns of all forms of written communications and their authorship". Raisig, in 1962, used the term `statistical bibliography' and defined it as "the assembling and interpretation of statistics relating to books and periodicals to demonstrate historical movements, to determine national and universal research use of books, and journals, and to ascertain in many local situation the general use of books, and journals". Raisig's definition of statistical bibliography is regarded today as one of the most classical definitions of bibliometrics. The British Standards Institution has defined Bibliometrics as "the study of the use of documents and patterns of publications in which mathematical and statistical methods have been applied". Sengupta (1988) defined bibliometrics as follows: "Organisation, classification, and quantitative evaluation of publication patterns of all macro- and micro-communications, along with their authorships by mathematical and statistical calculus". Though the term bibliometrics has a relatively recent origin its usage and practice, go as far back as to the second decade of the 20^6 century, either under the name Statistical Bibliography, a term coined by Wyndham Hulmeor without any name at all.

Three fundamental laws actually laid the solid foundation of bibliometrics. They are:

- Lotka's Inverse Square Law of Scientific Productivity (ef. *J. Washington Acad.* Sci. 16**,** 323, 1926);
- Zipf's Law in Linguistics (cf. *The psychobiology of language,* Houghton-Mifflin, Boston, 1935, USA); and,
- Bradford's Law of Scattering of Scientific Papers (cf *Documentation,* Washington D.C., Public Affairs Press, 1950, USA and *Engineering* **137, 85-86,** 1934).

The above three laws are respectively based on:

- the number of authors contributing in a discipline or other fields;
- the ranking of word-frequency in a particular set of documents; and,
- distribution of periodical publications in a defined area of knowledge or number of papers in a set of journals.

Alfred J. Lotka proposed his Inverse Square Law correlating contributors of scientific papers to their numbers of authors in a discipline or other field. He formulated a general equation for his law: the relation . . . . between the frequency "y" or persons making "x" contributions is:

$$x^2 \infty 1/(y)$$

$$\text{or } x^2y = C$$

where "y" is the frequency of authors contributing "x" papers each, and C is a constant.

George Kingsley Zipf's Law states that in a long textual matter, if the words are arranged in their decreasing order of frequency, then the rank of any given word of the text will be inversely proportional to the frequency of occurrence of the word. Mathematically Zipf's Law can be written as:

$$r \infty 1/ f$$

$$\text{or } rf = C$$

where, "r" denotes rank word, and "f" the frequency of occurrence of the word, and C is a constant.

S.C. Bradford enunciated his Law of Scattering as: "Articles of interest to a specialist must occur not only in the periodicals specialising in his subject, but also, from time to time, in other periodicals, which grow in number as the relation of their fields to that of his subject lessens and the number of articles on his subject in each periodical diminishes". Mathematically Bradford's law can be stated as:

$$f(x) = a + b\log(x)$$

where, f(x) is the cumulative number of references contained in the first "x" most productive journals, and "a", and "b" are two constants.

Bradford concluded that if scientific periodicals are listed in decreasing sequence of productivity of papers on a specific subject, then they stand divided into a nucleus of periodicals more particularly devoted to the subject, and several other zones containing the same number of articles as the nucleus, when the numbers of periodicals in the nucleus, and succeeding zones will be:

$$1 : N : N^2 : N^3 : N^4 :$$

Bradford also made a graphical representation of his data on Applied Geophysics and Lubrication, which is popularly known as "Bradford's Bibliograph". He plotted his bibliograph with cumulative number of references (R) against logarithm of the cumulated number of journals (log n). The graphical form of his data points out some ambiguity in his Law of Scattering. Mathematically his theoretical speculation and his graphical representation are not identical. His theory only corresponds to the linear portion of his bibliograph, and does not tally with the initial rising slope or droop of his bibliograph. The reason for this deviation may be due to lack of comprehensiveness of the bibliographies used by Bradford. It may also be that Bradford's law holds good only upto a certain point, and does not strictly follow the observed pattern of distribution of scientific literature.

An exponential growth of literature on bibliometrics has been noticed during the last six decades mostly based on these laws. Many standard statistical tests have been made either to substantiate, modify, extend, link or criticise these laws by subsequent researchers. As a consequence, number of other important new empirical laws, and theoretical models have been developed on the subject. Some of the new empirical laws are: Price's Square Root Law of Scientific Productivity (1963), Garfield's Law of Concentration (1971) and Sengupta's Law of Bibliometrics (1973a)

Price predicted that the number of elite in Science is small compared to the total number of scientists. In his Square Root Law, Price claims that any population of scientists of size N contains an effective elite of size√2. In other words, Price's Law states: "Half of the scientific papers are contributed by the square root of the total number of scientific authors".

Garfield in his Law of Concentration predicts: "A basic concentration of journals is the common core or nucleus of all fields".

Analyzing 98,780 citation data collected to rank periodicals of eight biomedical disciplines, namely, Medicine, Biochemistry, Microbiology, Physiology, Pharmacology, Genetics, Neuroscience, and Biophysics, Sengupta found that the data in general are in agreement with the findings of Bradford. In addition his data pertaining to each of these eight biomedical disciplines, lead to an interesting conclusion, extending the law of Bradford, and proposed a new Bibliometric Law, which states, "During phases of rapid and vigorous growth of knowledge in a scientific discipline, articles of interest to that discipline appear in increasing number in periodicals, distant from that field". In other words, Sengupta asserted that during phases of rapid growth of a scientific discipline, the small group of journals accounting for the larger part of the significant literature of the subject, contain relatively high proportions of unrelated journals. Mathematically, Sengupta's Law can be stated as:

$$f(x + y) = a + b\log(x + y)$$

where, f(x + y) is the cumulative number of references contained in the first (x + y) most productive journals, "x" is the number of most productive journals of the same field, and "y" is the number of most productive journals distant from the field, and y > x, and "a" and "b" are two constants.

Also, many bibliometric models have been designed to provide a sound conceptual understanding of these laws. Some of the pioneers of bibliometrics are: B. C. Brookes, F. F. Leimkuhler, W. Goffman, B. C. Vickery, M. G. Kendall, P. E. Cole, O. V. Groos, R. A. Fairthorne, S. Naranan, E. T. O'Neill, M. M. Kessler, H. G. Small, M. B. Line, M. P. Carpenter, F. Narin, H. H. Fussier, R. N. Broadus, K. Subramanyam, T. Radhakrishnan, 13. M. Hill, B. Mandelbrot, G. M. Dobrov, J. V lachy, A. Bookstein, A. E. Schoor, M. C. Drott, A: J. Meadows, D. B. Worthen, E. Willkins, F. Campbell, and others.

Recent trend shows that bibliometric literature covers nearly 25% of the, total contributions of library science. A sizeable proportion of bibliometric literature relate to citation analysis which is based on a hypothesis that any act of citing the author of an earlier paper is always meaningful.

As bibliometrics lies between the border areas of Social and Physical Sciences, its techniques have extensive applications equally in Sociological Studies of Science, Information Management, Librarianship, History of Science including Science Policy, Study of Science of Science and Scientists, and also in Social Science, and Social Scientists.

Some of the areas where bibliometric techniques are consistently being applied ar enumerated below:

- to identify research trends in different scientific disciplines;
- to quantify growth of research in different areas of knowledge;
- to estimate comprehensiveness of secondary periodicals;
- to identify users of different areas of knowledge;
- to identify authorship trends in documents of various subjects;
- to measure usefulness of ad-hoc and retrospective SDI services;
- to develop experimental models by correlating or bypassing the existing ones;
- to study past and present publication trends in any subject;
- . to forecast future publishing trends of any subject;
- to rank core periodicals of different scientific disciplines; .
- to formulate an effective need-based acquisition policy within the limited budgetary provisions, and that too without detriment to research interest of the parent organization;
- to initiate effective multi-level network systems;
- to adopt an effective weeding and stacking policy;
- to regulate inflow of information, and their judicious communication;
- to study obsolescence and dispersion of scientific literature (clustering, coupling, and mapping of scientific papers);
- to predict productivity of publishers, individual authors, organization, country, or that of an entire discipline;
- to design automatic language processing for auto-indexing, auto-abstracting, and auto-classification; and
- to develop norms for standardization.

9.5.3 : Informetrics

Informetrics is a term synonymous and analogous to bibliometrics, scientometrics, and librametrics. It is a concept introduced and practiced by FID (International Federation of Documentation). FID consider information as an important transactable commodity, a basic component for socio-economic development of a country. FID, therefore, in its 1980 May session, approved of the establishment of a *Committee on Informetrics,* designed it to deal with all mathematical aspects of information having an impact and relevance at international level FID wanted to assign the following tasks which have international relevance, to the Committee:

- to prepare international conference on informetrics;
- to lay down standards and definition for informetric research;
- to fit the U.D.C to research in informetrics;
- to lay down a detailed curriculum for the special education and training to

  information experts; and,
- to support the advancement of international cooperation in informetric research.

With these objectives in view, FID set up its first committee on informetrics (FID/ IM-Committee) in 1981 with its secretariat at the Institute fur Informetrie and Scientometrie at Bielefeld, West Germany, with Prof. Otto Nache as its first chairman. Besides Prof Nache, FID nominated some other experts from all over the world who are well known for their original contributions in scientometrics and other related disciplines as members of the first FID/IM Committee. Amongst these members, I.N. Sengupta was the FID nominated member representing the Asian countries. He sent a detailed note on Informetry to Prof. Nache, Director, Institut fur Informetrie and Scientometrie. A brief version of the note is reproduced in the following section.

"Informetrics is a well-practised measuring technique of Information Science. It is operative mostly through the techniques of statistical and mathematical calculus. Informetrics deals with all aspects of information. In the era of literature explosion, informetrics is considered as a vital commodity for generation of human knowledge, and this commodity is a transactable one. In addition to information handling, information storing, and information transferring at national and international level, it integrates and consolidates various mathematical and statistical measuring techniques to highlight the importance of information, its productivity, information theories, like, utility, cybernetics, stochastic, games, and also on decision-making. Further, information techniques are consistently applied to interpret various complex characteristics of information society. It also integrates various aspects of information technology, like, technologies related to media, computer, cybernetics, tele-communication, and networking.

"One of the major objectives of informetrics is to ensure complete informationalisation within human society. Informationalisation, on the other hand, helps to lead human society into a stage of complete industrialization, and thereby directly it works for the growth and upliftment of the socio-economic structure of human society. We may perhaps conclude saying that, informetrics based on statistical and mathematical calculus, play a key role for the growth and development of human society at national and international level.

There is not a single universally accepted definition of informetrics. However, we may define it "as a distinctive branch of information science which with the help of mathematical and statistical calculus elaborately studies:

- information on information,
- pertinence and relevance of information to society,
- efficiency and usefulness of national and international information systems, an(their services,
- role of formal and informal channels in scientific communications,
- efficiency of information products,

- soundness of information theories (e.g., stochastic utility, decision-making etc.),
- information technologies, like, media, telecommunication and networking,
- to determine intra and inter-disciplinary relation of information contained in the bibliographic references in macro- and micro-documents,
- citation habits of scientists,
- quantitative, and qualitative analysis of scientific journals, which contains most nascent information, and
- obsolescence, scattering and seepage of information".

M. Morales *(cf. Interat. Forum Int. Doc.* **10,** 15-21, 1983) has laid great importance to informetrics, and its applications. He views informetrics as a scientific activity relating to information, and thus an integral part of information science. All its metric aspects are mainly designed to improve the efficiency, and flexibility of information products, information establishment, while handling, storing, and transferring information. Morales identified some potential areas, where informetric studies can be very profitably undertaken:

- quantitative growth of literature,
- obsolescence and scattering of information,
- efficiency in information products and services, Science and Technology, and their overall production,
- efficiency of the information system, and information establishment in general,
- the role of different kinds of documents as means of scientific communications,
- information pertinence and relevance,
- ranking of periodicals and other serials by various parameters,
- the role of informal channels in scientific communications,
- overlapping of subject contents between periodicals, and serials,
- citation habits of scientists and the growing role of citation analysis, and

intra-disciplinary and inter-disciplinary relations determined on the basis of bibliographical references.

9.5.4 : Scientometrics: Its Genesis, Scope, Definition and

Scientometrics is one of the four measuring techniques of Library and Information science. It is a concept analogous to bibliometrics, and informetrics. It is a new emerging area of research. It uses various mathematical and statistical techniques for the evaluation, and advancement of Science, and its level of development, impact, and relevance to the human society.

The concept of scientometrics was an active research area in USSR, and in the eastern European countries. Scientometrics, from its early analytical beginnings has grown into an active branch of Science of Science. It is considered as a quantitative technique to measure various parameters of science at individual, institutional, national, and even international level.

Thus, *Scientometrics,* may be defined as:

"A quantitative, and qualitative measuring technique for evaluation, and interpretation of science of science including its different activities like: productivity, progress, organisation, and management of Science, by application of mathematical and statistical calculus."

Noted Russian scientometrists Nalimov, and Mulchenko in their book (cf. *Scientometrics: A study of science as an information process, Moscow,1969, p.-12),* defined scientometrics as:

"A quantitative method of investigating the development of science as an information process".

B.Ya. Brusilovsky in his book (cf. *Mathematical models in forecasting Science organisation, p.-6,* Kiev, 1975) defined scientometrics: as a set of mathematically correct methods in science studies".

S.D. Haitun (cf. *Scientometrics, 2, 63-85, 1980),* defines scientometrics:

"as an approach the science of science aimed at a reproducible measurement of Science which reveals its objective in quantitative, regularities".

According to J.P. Courtial (cf. A co-word analysis of scientometrics. *Scientometrics,* **31,** 251-260, *1994):*

"Scientometrics is a hybrid field made of invisible college and a lot of user, thus controlled by both scientific research, and final uses".

Sengupta (cf. Growth of Biophysical literature. *Scientometrics, **8,** 365-375,1985)* recently outlined the objective of scientometrics as:

"To evaluate quantitatively recent growth of any basic scientific discipline, and the factors responsible for the steady growth in research activity in that area of knowledge in the post-war period".

He further stresses that the quantitative evaluation of Science along with the inter, and intra comparison of scientific activities, including its productivity, and progress may be of great help to the Science administrators in planning proper utilization of available financial, and manpower resources more economically for the betterment of socio-economic structure of a country. One of the basic aims of scientometry is to measure the organizational and management aspects of Science.

It may be noted that when any branch of knowledge starts growing, learned societies are formed to cope with the growth of knowledge of that field. Subsequently specialist periodicals solely restricted to that discipline start appearing to publish the findings of the research in that field. Prior to it, the researchers may publish their results in periodicals of a related field. *Scientometrics* started appearing from Hungary, launched by the Elsevier Scientific Publishing Co., Amsterdam in 1978 in collaboration with Akademai-Kiado, Budapest. Incidentally, *Scientometrics* is the first journal on the subject with an international Advisory Board of Editors, including Dr. I. N. Sengupta from India.

In *Scientometrics,* vol-1, ***no.* 1,** 1978, some of the members of the editorial board wrote about the aims of scientometrics. J. Vlachy stated that the aim of scientometric research is the discovery of the equations, which subsist between the elements of phenomena, was a program of one group of scholars, men have to be weighed, not counted, warned some humanists. H.T. Beck felt that "Scientometrics may help in the more economical and balanced utilisation of the available funds, thus increasing the efficiency of research." He warned that, "both the over-estimation, and underrating of its potential should be avoided ... scientometrics is neither an omnipotent diagnostic tool, nor a panacea, it can be most instrumental in revealing, and solving a variety of problems in Science".

G.M. Dobrov (cf. *Scientometrics,* 1, 1-2, 1978), mentioned in his editorial statement that, "The greatest emphasis and longest historical past, are associated with scientometric work on 'analysis of the informational parameters of Science development, such as the number of papers, patents, journals, laws of `ageing', and `dissipation' of scientific information, structure of the flow of scientific documents, citation processes, etc.". To elaborate the problems of scientometric studies, Dobrov feels that scientometric investigation should be aimed at determining series of dynamic indexes describing science systems in the process of their development taking into account that science is a probability system, consequently the results obtained are probabilistic in nature.

He recommended that "Emphasis (in scientometric studies), should be placed on the necessity of mathematical analysis of the empirical data of scientometrics with special reference to the application of statistical methods, graph theory, cluster analysis etc.". He further opined that, "the quantitative methods of scientometrics should always be followed by a qualitative analysis of the results". The efficiency of the application of scientometric results in practical Science management strongly depends on the correct interpretation of the data ... scientometric investigation should be aimed at determining sciences.

Derek de Solla Price (cf. Editorial statement. *Scientometrics, 1,* 2-3, 1978), asserted that the major aim of scientometrics is to do scientific analysis of science mathematically. He said, "however successful we are in understanding the productivity of scientists, and the mechanics and pattern of coherence of scientific creativity, we shall still need historians, sociologists, and psychologists of Science, for those types of analyses that cannot be expressed in metric terms." He further stressed that, "since science and scientific activity is peculiarly measurable, and peculiarly regular in its behaviour, even compared with the models of scholarship".

It may be mentioned that the quantitative evaluation in Scientometrics should always be backed by qualitative analysis. The success of scientometric studies for management of science depends on the comprehensiveness of data collected, and on their accurate interpretation. *Scientometrics* generally publishes original investigations on quantfiable features, characterizing the nature and mode of development of science. It aims to provide an international forum for publication of original studies, primarily reports, communications, review papers, letter to the editors, reviews, information about international meetings, and events on quantitative aspects of science of science policy. Many of the techniques of bibliometrics to quantify the development and progress of Science, relevance to and impact of, Science in society.

9.6 : Summary

This unit discussed:

- analytical studies on library related data;
- citation analysis: its merits and demerits;
- new method of citation analysis;
- elimination of bias due to country of origin and language of the texts of the source journals;
- citation corrections for newly incepted periodicals;
- role of three bibliometric parameters in citation analysis;
- Garfield's cito-analytical products;
- four analytical measuring techniques in Library Science;
- librametric analysis for library-related data;
- bibliometric analysis for library-related data;
- informetric analysis for Iibrary-related data; and
- scientometric analysis for library-related data.

10.0 : Objectives

After reading this Unit you will understand:

- the various components of user studies;
- why such studies are conducted;
- how to conduct such studies; and

the importance of these studies in the field of informetrics and scientometrics

10.1 : Introduction

Of the Five Laws of Library Science enunciated by Ranganathan the second, thirc and the fourth laws. respectively are `Every Reader his book', `Every book it: reader' and `Save the time of the reader'. Readers are users of the library. Three laws out of five devoted to reader clearly underline the importance of users it library and information services. They are equally important in the field of scientometric: and informetrics as well. Hence, user studies form an important area in informetric: and scientometrics in as much as these studies generate important indicators as to the types of documents users use heavily, moderately or scarcely; the various types of services they need, and the urgency with which they need these services; the reasons why particular documents and services available in the library are not being used optimally; the satisfaction level of the users; and so on. User studies also bring into the focus the various categories of users of a particular library or information centre.

10.2 : User Studies

User studies are basically concerned with the users. Analysing the results of these studies we come to know about the extent of use of the various types of documents and services. In the **use studies** we are mainly concerned with the use of the documents and services and not with the users. For example, conducting citation studies we find out the extent of use of the various types of, documents knowing little about the types of users making use of those documents.

TOP

10.2.1 : Definition

User - In LIS parlance, a user is a person who makes use of documents or services of a library or information centre. For example, in a university library, the users are students, faculty members, administrative and other staff. In a scientific library, mostly scientists are the users.

Study - Here, study means a thorough examination of a subject conducting surveys, examinations, and so on.

User Study - Now, combining the two definitions given above we derive the definition of user study as follows. It is the thorough examination of the users of a library or information centre in terms of their types, the categories of documents and services they use, the urgency with which they need the documents and services, etc., using one or more of the approved methods

10.2.2 : Purpose

The purpose of user studies are:

i) to identify the various types of users of a particular library or information centre;
ii) to identify the types of users of an information service or product;
iii) to identify the types of documents used heavily, moderately or scarcely and the reasons thereof;
iv) to identify information services and products not at all being used and the reasons thereof;
v) to measure the satisfaction level of the users;
vi) to detect the strengths and weaknesses of various collections;
vii) to frame the collection development policy based on facts; and
viii) to develop the collection in a scientific way.

[EDIT]

10.3 : Methodology

Various methods are employed to conduct user studies. Some of the well-known methods are: Questionnaire method, Interview method, Dairy method, Observation method, and Library record analysis. We shall now briefly discuss these methods one by one.

10.3.1 : Questionnaire Method

This particular method is most heavily used to elicit required information from the users because of several advantages. The questionnaire can comprise both close-ended and open-ended questions. In the close-ended questions, the respondents are simply to tick the most appropriate item recorded in the questionnaire.

For example, a close-ended question may be like this:

How often do you use the library?

Everyday ☐ Quite Often ☐
Sometimes ☐ Never ☐

In this question the respondent will tick the item that is most appropriate *for* him. *An* open-ended question may be like this:

Are you satisfied with the services being provided by the library? If not, give reasons.

Here, the respondent will describe in his own words the reasons for his dissatisfaction. This method can encompass a huge number of respondents spread over a wide area, such as a state, country, region or even the world. The questionnaire can be distributed using e- mail or postal services. The reminder can also be sent using the same mode. In most cases, there is no need for the investigator to personally meet or talk to the respondent. These factors reduce the cost of eliciting answers very heavily. Moreover, this method helps a respondent to fill in the questionnaire at leisure and express his opinion freely, which he may not do in a face-to-face interview for some reasons or the other. One of the major disadvantages of the method is the response rate, which at times may be as low as 30% and the data may not reflect the actual state of affairs. If this problem can be overcome, then this is really a very good method for data collection.

10.3.2 : Interview Method

This particular method of data collection is very useful where the respondents are by and large local and can be approached easily either through personal visit or through telephone. Take, for example, a school or a college library. In this case all the the respondents forming the sample are available within the campus. The investigator can approach all of them, make appointments, and can interview one by one

The interview can be taken in any one of the four styles described below.

Structured - In this, the questions and the answer categories are determined in advance by the investigator.

Standard open-ended - In this, the questions are decided in advance by the investigator but the respondent may answer the questions in his/her own way.

Interview guide approach - In this, the topics to be discussed and the time of meeting are decided first by the investigator in consultation with the respondent. During the meeting, the investigator questions the respondent in his own way and the latter replies also in his own way.

Unstructured - In this, except the topic, nothing else is fixed. This method ensures very high response rate in as much as the investigator can meet the respondent personally and elicit the required information through interview. The accuracy of the replies is also high as the respondent can seek clarification on any of the questions

10.3.3 : Diary Method

In this case the investigator requests the users of a library to record in his dairy all such activities as are related to the study being conducted. The user is to record each activity he performs during each of his/her visits to the library during a fixed span of time, which may be one week or more. For example, a scientist may be asked to record the titles of the journals, books and other documents he/she had consulted in the library over a certain period of time. After the period is over, all the diaries are collected and the data recorded are checked. If necessary clarification may be sought from the respondent, and finally the data are analyzed to generate various indicators

10.3.4 : Observation Method

Scientists have for long used this method for, the collection of data. Nowadays, this method is widely used in Social Sciences and other fields as well.

Usually, four types of observation are undertaken by investigators. They are: i) Participant observation, ii) Non-participant observation, iii) Non-controlled observation, and iv) Controlled observation.

Participant observation - In participant observation, the observer actively participates with the activities of the group under study without making other members know that their activities are being closely observed by someone. This makes the members act freely and all their activities can be watched under natural condition.

Non-participant observation - In this case, the observer does not take part in the activities of the group and simply observes them from a distance. Absolute non-participant observation is not desirable as it makes both the parties uncomfortable. Hence, the observer keeps himself busy with some activities of the group and keeps watch on the activities of other members.

Non-controlled observation - Here, observations are made in natural surroundings, and the activities are performed without being influenced and guided by any external force. In this case, the observer visits the place of the occurrence of the phenomenon to study the situation. The method has a few drawbacks such as that the observer may be biased, he may generalize the situation from a few stray incidents; interpret the data based on his/her limited knowledge, etc. For example, an audit party while visiting a technical library observed twelve copies of a book lying on the shelf and raised an audit objection as to why so many copies of the book were purchased leading to unnecessary wastage of public money. The fact was that the institute to which the library was attached was conducting several short-term courses in a year on the subject treated in the book, and each participant to the course was given a copy to enable him/her follow the lectures properly. The books used to be kept on the shelf when there was no course. Of course, such cases of wrong interpretation are not many. If the observer gets the data verified, the cases of wrong interpretation will be much less.

Controlled observation - To overcome the limitations of the non-controlled observation, controlled observation technique has been devised. In this case, control is exercised on the phenomenon as well as on the observation. For example, in the children's section of a library, the behaviour of the same set of children may be observed for days together to study their reading habit. The phenomenon can be observed by a teacher, a student, or by an electronic device like video camera. This technique affords greater precision and objectivity, and can be carried out time and again under similar conditions.

10.3.5 : Library Records Analysis

Many libraries maintain records relating to the acquisition of books, periodicals, and other documents; circulation of books; inter-library loan; document delivery; queries handled; and so on. Analysis of these records may provide many useful indicators, such as, the rate of growth of the collection; subject-wise growth of the collection; heavily used books and periodicals; strong collection; weak collection; and so on. This method cannot take into account the use of books and other materials within the library premises as they are not getting issued, and hence there is no record.

10.4 : Planning a Survey

Planning is one of the most important factors in conducting a survey. If planning is correctly done, the objectives of the survey can be achieved to a very great extent. Bad planning sometimes vitiates from achieving basic objectives of the study.

10.4.1 : Basic objectives of the Survey

The first step in planning the survey is the defining of the basic objectives in clear terms. Once this is done, deciding the methodology for the survey becomes easier. Moreover, the persons who will be involved in the survey, i.e., the staff and the respondents, may be informed about the objectives for seeking their cooperation. Suppose, the objective of your survey is the determination of the users' needs, then this can be communicated to the users as well as the staff who will be involved in the survey.

10.4.2 : Scope of the Survey

The next step is to determine the scope. Defining the scope will include apart from the users, the place, time, types of documents and services, etc. There are various categories of users in a library. In a college library, there are teachers, students; and the administrative staff. All may be users. Are you going to consider all the users, or only a particular category or categories for your study? You may consider teachers, students, or both. The next point to consider is the place. It can include the entire country, a state, district, city, or simply the campus of the institution. The third point is the time. You can observe the use of books by students for a week, a fortnight, a month, or for a longer period. The next point is the type of documents. You may consider in your study only the use. of books or periodicals or both. Non-book materials can also be the focus of your study. Service is another factor to consider. You can include in the survey the Internet facility provided to find out the extent of its use.

10.4.3 : Methodology for Data Collection

There are several methods for data collection and one that suits the purpose should be selected. Two or more methods may be combined, if necessary. For example, the interview method may be combined with the questionnaire method. Let us say that taking into consideration the cost, labour, pace and extensiveness of data collection, you have opted for the questionnaire method for surveying the needs of the students and teachers of a college library. There are distinctly two categories of users here. Hence, two different questionnaires are required. One combined questionnaire may also serve the purpose, but it may be lengthy and there may be questions which may be asked from one group but not the other and vice versa. Now, let us see how the two questionnaires can be framed.

10.4.4 : Framing of the Questionnaire

As this is a user study, the questionnaire should start with such personal details as Name, Age, Sex, Designation, Class/Year (applicable for a student), present address, permanent address, etc. In use studies such details are usually not required. Sometimes the name of the person may not be asked for if frank expression of opinion is sought.

A user of a library usually needs books, periodicals, and other documents, and services, such as, photocopies of documents, assistance in consulting the catalogue, guidance relating to the location of various collections in the library, career guidance, etc. With the advent of computers, many libraries have computerized their catalogues and got the Internet connection. As such the need for assistance in the use of computerized catalogue as well as the Internet facility has become much more obvious. These are some of the expressed needs. There may be some implied needs. For example, there may be shortage of space in the reading room or quite often there may be long queues at the circulation counter. Every user needs a comfortable seat in the library for reference and study without any disturbance. He/ she also would want a book to be issued with the least delay. In user studies we may also wish to find out as to what extent the needs of users are satisfied.

With these points in mind, the questionnaire for students may be framed as shown below.

QUESTIONNAIRE

Note: This questionnaire is being used to ascertain your actual need as to the use of the library. Please fill in the questionnaire without any hesitation so that we can know your difficulties and problems in the use of the library. Once we know your problems and difficulties, we can try to solve them and make this library really useful for you. All your opinions are welcome and will be kept confidential. Writing of name or signing on the questionnaire is optional. In a multiple-choice question, please tick the item that is most appropriate in your opinion

) Name (optional)

2) (a) Age (b) Sex...Male/Female

3) In which Class/ Year are you studying?

4) (a) Which Course of Study you are pursuing ...Arts/Science/Commerce

(b) Pass/Honours

5) Present address ....

6) Permanent address ....

7) Do you use the library for books? Yes/No

8) If yes, do you use them in the library? Yes/No

9) Are you permitted to borrow the books for home reading? Yes/No 9).

10) If yes, how many books can you borrow at a time?

11) Is the number of books allowed to be borrowed at a time sufficient? Yes/No

12) If no, how many books would you like to borrow at a time? Please

specify......................

13) For how many days can you retain a book with you?

14) Is the period sufficient? Yes/No

15) If no, for how many days do you want to retain a book with you? Please

specify...........

16) Do you get the book easily? Yes/No If no, state reasons

……………………………………………………………………………………………………

……………………………………………………………………………………………………

17) Apart from textbooks, do you use other type of books, e.g. dictionaries, encyclopedias, fiction, essays, poems, short stories? Yes/No

18) If yes, please specify.

19) How often you do not get the required book at all? Please specify........................

20) Do you read periodicals? Yes/No

21) If yes, please name the periodicals you read most.

……………………………………………………………………………………………………

……………………………………………………………………………………………………

22) Do you want some more magazines to be added in the collection? Yes/No

23) If yes, name the magazines.

……………………………………………………………………………………………………

……………………………………………………………………………………………………

23) Do you read newspapers? Yes/No

24) If yes, please name the newspapers you read most

……………………………………………………………………………………………………

……………………………………………………………………………………………………

25) Do you read any other type of documents? Yes/No

26) If yes, please specify the types of documents you use.

……………………………………………………………………………………………………….

……………………………………………………………………………………………………….

27) Does your library provide inter-library loan service for students? Yes/No

28) If no, do you feel such service will be helpful to you? Yes/No

29) Does your library provide photocopy service? Yes/No

30) If no, do you feel such service is necessary for you? Yes/No

31) Do you need sometimes the assistance of the library staff in the consultation of the catalogue? Yes/No

32) If yes, do, you get the assistance easily? Yes/No

33) If no, state the reasons

……………………………………………………………………………………………………….

……………………………………………………………………………………………………….

34) Do you know the locations of each and every collection in your library? Yes/No

35) If no, please state how do you locate the collection?

- By trial and error method
- With the help of a friend
- With the help of a library staff

36) Does your library provide career guidance? Yes/No

37) If no, do you think such a service will be very helpful to you? Yes/No

38) If you fail to locate a document in the library, or a piece of information in adocument, does the library staff help you? Yes/No

39) If no, do you feel such a help is very essential? Yes/No

40) Are you aware of usefulness of Internet for your study? Yes/No

41) If yes, do you feel installation of Internet facility in your library will be highly useful for you? Yes/No

42) How often do you visit the library?

- Everyday
- Several times a week
- Several times a fortnight
- Several times a month
- Rarely

43) Do you get a comfortable seat in the reading room whenever you visit the library? Yes/No

44) If no, how long you are to wait to get a seat?

- Less than 15 minutes
- More than 15 minutes
- More than half an hour

45) How long does it take to get a book issued?

- Less than 15 minutes
- More than 15 minutes
- More than-half an hour

46) State the level of your satisfaction as to your use of the library

- Fully satisfied
- Highly satisfied
- Satisfied
- Moderately satisfied
- Not satisfied

The questionnaire can be used for teachers as well by making some changes in the personal details and questions relating to some more services such as bibliography compilation, current awareness service, selective dissemination of information service, etc.

10.4.5 : Testing of the Questionnaire

The above questionnaire for college students before distributing copies to the students for eliciting answers it is necessary to test the questionnaire to ensure that; (i) the students understand every question clearly, (ii) no relevant question has been left out, (iii) and the questionnaire is not very lengthy. The investigator may also observe the time taken by the test students to complete the questionnaire, so as to work out the average time needed to fill in the questionnaire.

The steps for the testing of the Questionnaire are as follows:

1) **Selection of the Sample -** The sample of respondents of various categories should be more or less of the same proportion as of the original. Suppose, it is decided to take forty students from each class pursuing Pass Course and twenty students from each class pursuing Honours Course for your survey. In that case, if ten students are selected from each class of the Pass Courses, then five students from each class of the Honours Courses should be selected to maintain the same proportion.

2) **Distribution of the Questionnaire -** The distribution of the Questionnaire to the sample (population) may be preceded by a brief talk explaining them the purpose. The respondents comprising the sample should be encouraged to ask for clarification if they did not understand any particular question.

3) **Checking the Time Required to Fill in the Questionnaire** The time should be reckoned from the distribution of the questionnaire to the receipt of the last filled-in questionnaire. From this an approximate idea can be had of the time required for filling in the questionnaire. If a questionnaire is too lengthy and requires substantial time to fill in, respondents will not be interested to fill in the questionnaire.

4) **Checking the Filled-in Questionnaire -** All the filled-in questionnaires should be checked to see whether the respondents have understood each question correctly. If several respondents answered the same question wrongly, then the question should be checked and ambiguity, if any, should be removed. If there is an uncommon word or term in the question, it should be replaced with a common one.

There may be some questions left out completely unanswered. If these questions are consistently toward the end of the questionnaire, then it may be reasonable to conclude that it was due to lack of time. If this is not the case, then the reasons should be found out and the questions should be modified.

Finalization of the Questionnaire - Taking all the factors into account, the questionnaire should be finalized and the required number of copies made for the distribution to the sample population of students.

10.4.6 : Selection of Sample

In a college library, the teachers and students as they are the main users of the library. If the number of teachers is not very large, say, below fifty then all the teachers can form the sample. Usually, the number of students in a college is large. Hence, only a part of the students from each class is to be chosen for the survey. Suppose, there are fifty students in a class and only ten are to be chosen in such a way that different categories are included in that. One can start from the left, and select every fifth student, and that will form the sample. Instead of left, one can start from the right, or from the end, and so on. This type of sampling is called stratified sampling and is useful where the population is heterogeneous.

In sampling, the size of the sample is of paramount importance. If the sample is very large, conducting the survey becomes time consuming, laborious, and costly. It is not the case that with a very large sample, you get better results. The larger the population, the smaller can be the sample in terms of proportion. For example if the population size is 100, you can have forty or fifty respondents in your sample If the population size is 10,000 you can take just 500 (that is, 5 per cent) respondent for the sample, and that will be adequate

10.4.7 : Distribution of the Questionnaires

If the respondents are located within easily accessible distance, the questionnaire; can be distributed personally or through a courier. If they are dispersed over a wide area, say , a district, state, or a country, then one may have to utilize postal services. It may even be possible for the students to access the questionnaire via a network connection - intranet or the Internet. Always, indicate at the beginning of the questionnaire and/or the covering letter a deadline by which the questionnaire is to be returned. Two-weeks or one month time may be given to return the questionnaire

10.4.8 : Receipt of the Questionnaires

If the respondents are located nearby, follow up operation can start just after the expiry of a week. The respondents may be phoned up or personally met to request the return of the questionnaire at an early date. In the case of mailed questionnaires, the respondents may be reminded fifteen days before the deadline. If the number of returns is less than thirty per cent, vigorous efforts are called for to get more returns, otherwise the data may not be sufficient to project the correct scenario.

10.6 : Analysis of Data

Data can be analyzed in a number of ways. If we collect data using the Questionnaire given above taking college students as sample, then we shall be able to analyse the data in the following ways:

1. According to Age

We can use the data to determine the average age of the students of the first year, second year and third year. We can also have a breakdown according to sex (male, female) and also according to the course of study, i.e. BA,

2. According to Course of Study

Using this data, we can find out the number of students in each course, i.e. BA (Pass Course), BA (Honours Course), etc. and can have the breakdown according to sex.

3. According to Place of Residence

Using the present address, it is possible to find out how many students are residing in hostels, and how many are coming from outside. We can also have a breakup according to courses of study and sex. Using the permanent address, data can be generated as to the number of students coming from various states, and from within the city and the state, we can also have a breakup according to courses of study and sex.

4. **Use of Books**

Analysis of this data will give us information as to ;

i) the use of the library for books by students with male, female and course-wise breakup,
ii) the number of books,;, student can borrow and the number of days the books can be retained,
iii) whether the number of books allowed to be borrowed, and the number of days they are allowed to be retained is sufficient,
iv) the type of books the students normally use,
v) and the frequency of not getting a book.

5) Use of Other Documents

This data yields information as to the use of periodicals, and other documents by students with sex-wise and course-wise breakup.

6) Library Services

Analysis of this data will indicate the types of services such as lending, inter-library loan, photocopy, reference, career guidance, and Internet services being provided by the library, and whether the students need such services.

7) Reading Room

Whether or not a student gets a comfortable seat in the Reading Room easily can be ascertained from the analysis.

8) Waiting Time to Borrow a Document

Long queues at the Issue Counter eats away valuable time of a student. How quickly a student can get a book issued can also be ascertained through the analysis.

9) Satisfaction

Indicators as to the satisfaction of a student with the library service can be generated using this data.

10.7 : Presentation of Results

The results of the survey are described in words accompanied by tables and diagrams. The text provides the summary of the results, the tables the detailed data, and the diagrams make the presentation attractive, easily understandable. We discuss here the tabular and the graphical presentations.

10.7.1 : Tabular Presentation

The tabular form is the most common form of presentation of results. The tabular form presents data of a survey which is not usually available in other forms. For example, in Table 3, you get the exact number of students in each course which one may not get from bar charts or pie diagrams. However, attention should be paid to the following basic qualities of a good table.

1) **Attractiveness:** In the table, the title, captions, stubs, figures, etc. should be cleanly typed with proper spacing of each cell. Use of gridlines for columns and rows increases the legibility of the table and adds to its elegance. To the extent possible, a table should be simple, and devoid of over-writings and hand-written corrections.

2) **Size:** A table should be of manageable size, and this point should always be borne in mind. The number of columns and rows in a table should be just adequate. How the size of a table may be reduced. Here are some guidelines:

Suppose, a study shows that the users of a Science library have used one thousand periodicals during the period of a survey with the frequency ranging from one to more than a hundred. If the frequency of use of every periodical is to be shown, the table will require one thousand rows, occupying about 20 pages in a journal. This is unacceptable. There may be some 30 periodicals whose frequency of use may be more than ten. Show them all in the table. For the rest, just mention the number of periodicals against a given frequency. An example is given below.

Table 5: Ranked List of Periodicals Used in Ramkrishna Science Centre Library during March 2000

Rank	Name of the Periodical	Frequency of Use
1	*****************	123
2	*****************	115
3	*****************	107
4	*****************	101
5	*****************	88
6	*****************	73
7	*****************	67
8	*****************	54
9	*****************	48
10	*****************	33
11	*****************	27
l2	*****************	21
13	*****************	19
14	*****************	16
15	*****************	14
16	*****************	14

17	*****************	13
18	*****************	13
19	*****************	12
20	*****************	12
21	*****************	12
22	*****************	1_1
23	*****************	11
24	*****************	11
25	*****************	11
26	*****************	10
27	*****************	10
28	*****************	10
29	*****************	10
30	*****************	10
31-49	19 periodicals	9
50 - 87	38 periodicals	8
88-140	53 periodicals	7
141 -210	70 periodicals	6
211-308	98 periodicals	5
309-441	133 periodicals	4
442-593	152 periodicals	3
594 - 769	176 periodicals	2
770-1000	231 periodicals	1

3) **Comparability:** Comparability is an important feature of a table. To ensure comparability, the data in the table should be placed in such a way that they are easily comparable. For example, in Table 3, we can easily compare the number of Pass Course students with the number of Honours Course students in **:** i) each course in one year or, ii) year by year. We can also compare the same course by course. The number of Pass Course or Honours Course students only year by year or course by course can also be compared easily.

4) **Clarity:** In most cases the target users of the tables are laymen though they are prepared by the expert statisticians. Hence, the content of the table should be understandable to a layman.

5) **Suitable for the Purpose:** The table should serve the purpose for which it has been prepared. It should neither include anything redundant, nor miss anything relevant.

6) **Correctness of Data**: All the data given in the table including the totals should be correct in all respects. In case, there is doubt about the correctness of some data, footnotes to that effect should be provided.

7) **Construction**: It is to be constructed in a scientific way as discussed here.

10.7.2 : Graphical Presentation

Data can be presented graphically in a number of ways. Some are mentioned below:

1) Unidimensional Diagrams:
   i) Line diagram
   ii) Simple Bar Diagram
   iii) Multiple Bar Diagram
   iv) Subdivided Bar Diagram
   v) Deviation Bar Diagram
   vi) Sliding Bar Diagram
   vii) Pyramid Diagram

2) Two-dimensional or Area Diagrams:
   i) Rectangles
   ii) Squares
   iii) Circles
   iv) Pie

3) Three-dimensional or Volume Diagrams:
   i) Cubes
   ii) Cuboids
   iii) Cylinder Blocks
   iv) Pyramids

As there are various parts of a table, similarly there are various parts of a diagram. Every diagram should be numbered and given a title. In research papers and many books, the number and the title appear below the diagram. To indicate a diagram, usually we use the abbreviation Fig. before the number such as Fig. 1 - Growth of Student Population of Ramlal College during 1990's.

Some of the basic qualities of a diagram are as follows:

1) A diagram should be neat, well-proportioned, attractive, and pleasing to the eye.

2) It should be accurate from the geometrical point of view. Suppose, the student population of Ramlal College is double compared to that of the Shyamlal College. If the student population is being represented by bars, then the size of the bar of Ramlal College should be exactly double the size of the bar of the Shyamlal College.

3) The size of the diagram should be decided keeping in view the size of the page and easy readability. It should neither be too big nor too small.

4) As there are many types of diagrams, the selection of the right type of diagram is important for presenting given data.

5) The diagrams may be vertical or horizontal. However, vertical diagrams are preferred as they are easy to follow.

6) The diagram must include the tags, and the scales so that the actual figures may be derived from them.

7) Unlike tables, we use various colours in diagrams to make it more attractive and easily understandable.

8) To extent possible, a diagram should be simple. Too much detail that complicate the comprehensibility of a diagram should be avoided. A diagram may be split to make it simpler.

Unidimensional Diagrams

1) **Line Diagram:** In a line diagram, only lines are drawn to represent the data. The lines may be vertical or horizontal. Compared to bars, lines occupy much less space. When the data to be presented within the bounds of a page is large, the line diagram is preferred to the bar diagram. The line diagrams given at 1B and IC are finding greater acceptance than the one at Fig. IA with the advent of software packages like Excel, SPSS, etc.

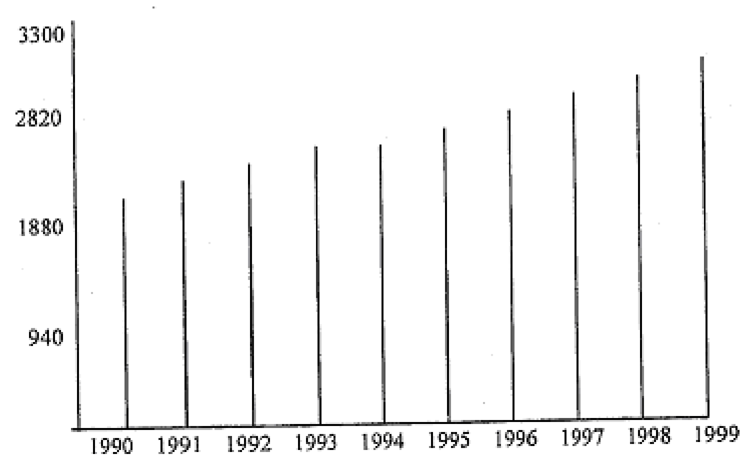
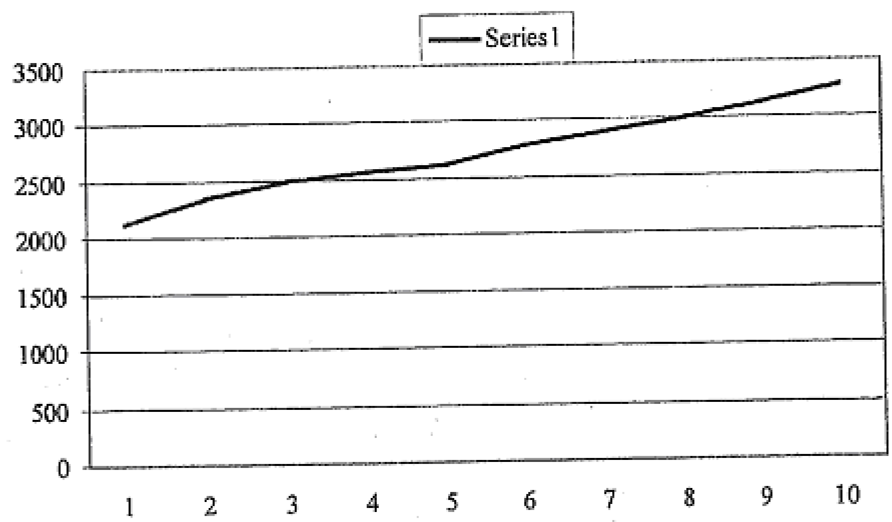
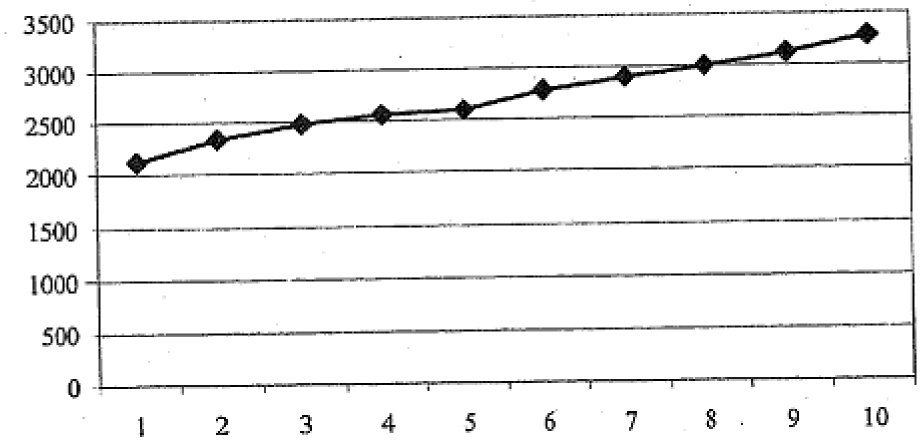


Fig. 1A: Growth of Student Population of Ramlal College during 1990's



1 to 10 in the diagram represent years 1990 to 1999 respectively

Fig. 1B: Growth of Student Population of Randal College during 1990's

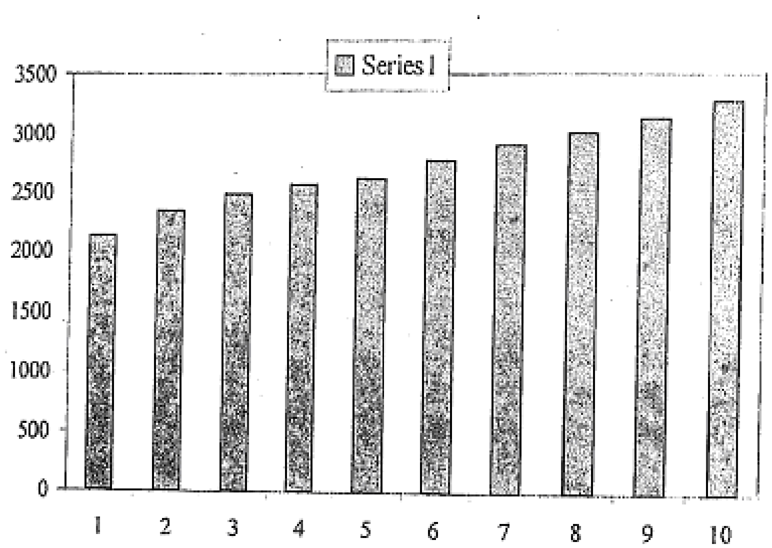


1 to 10 in the diagram represent years 1990 to 1999 respectively

Fig. 1C: Growth of Student Population of Ramlal College during 1990's

1) **Bar Diagrams:** A Bar diagram in reality a two-dimensional diagram. Here, we are considering them under uni-dimensional diagrams because only one dimension, i.e., the length, is being taken into account. In two dimensional diagrams we take into account the area which involves both the length and the breadth.

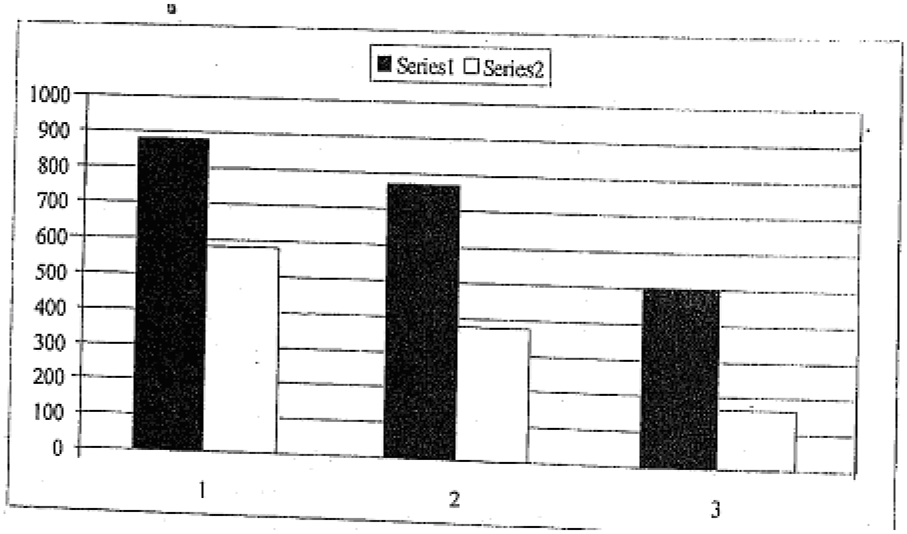
Simple Bar Diagram: Bar diagrams are quite similar to line diagrams. Here the lines are simply replaced with bars. Using the same data used to draw the line diagram in Figs. IA, 1B, 1C we can draw a bar diagram as shown below:



1 to 10 in the diagram represent years 1990 to 1999 respectively

Fig. 2: Growth of Student Population of Ramlal College during 1990's

Multiple Bar Diagram: Now, we present a multiple bar diagram with the data of Pass Course and Honours Course students of Ramlal College for 1999.



1 = BA students, 2 = B Com students, 3 = B.Sc. students
Long bar = Pass students Short bar = Hons. Students

Fig. 3: Pass and Hons. Course Students of Ramlal College during 1999

Subdivided Bar Diagram: Now, we shall draw the Subdivided Bar Diagram using the same data as used in Fig. 2

In the X axis, 1, 2, and 3 represent BA, B Corn and B Sc students respectively. The lower portion of each bar indicates Pass Course students and the upper portion Hons. Students.

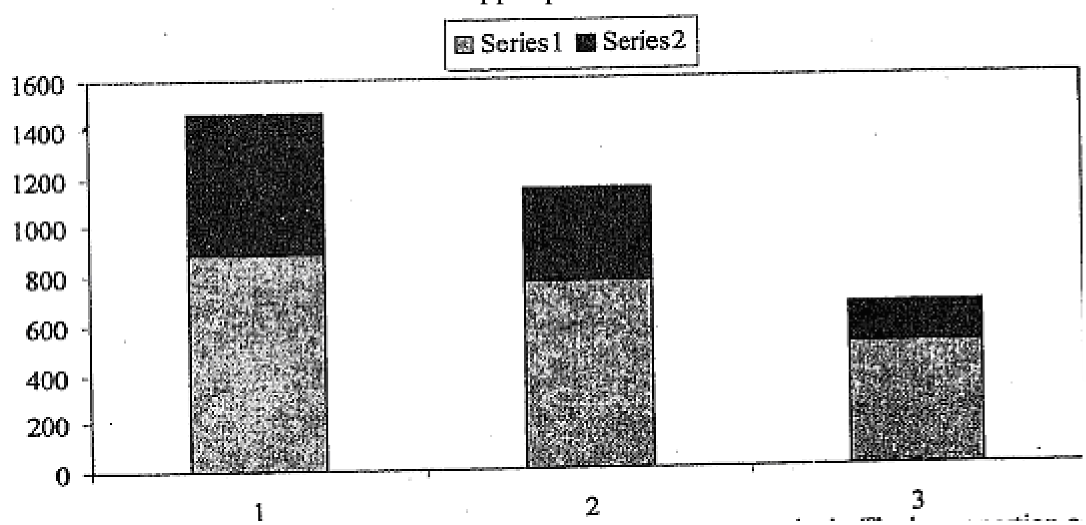


Fig. 4: Pass and Hons. Course Students of Ramlal College during 1999

Deviation Bar Diagram: The Deviation Bar Diagram is used to show the deviation from a given number. Suppose, Ramlal College Library subscribed the following number of periodicals during 1990 - 1998 as given in Table 5.

Table 5: Change in the number of Periodicals subscribed by the Ramlal College Library during 1990 - 1998.

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998
Periodicals	940	955	975	990	1005	1015	1040	1050	1080

The figure given below shows the deviation from the number 1000 in both the sides of the base line. In this figure, the base line is vertical. The base line can be drawn horizontally also.

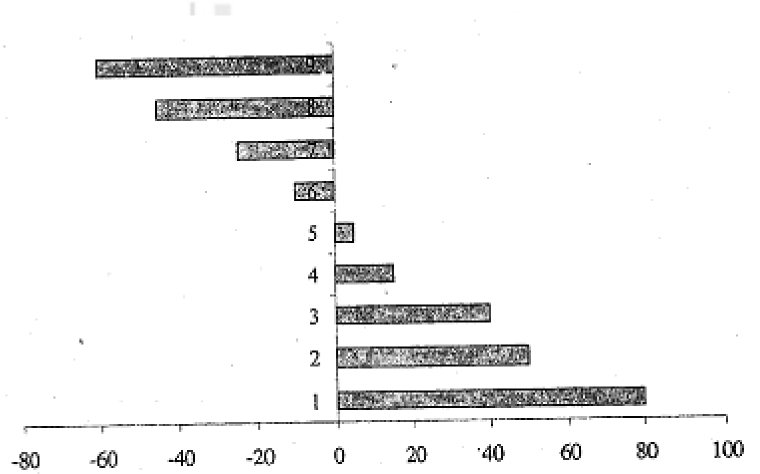


Fig. 5: Change in the number of Periodicals subsdcribed by Ramlal College Library during 1990's

Sliding Bar Diagram: Sliding bars are used to show the ratio of change in two parts of the unit. The total of the unit is always 100. Suppose, the male female breakup of the BA students of Ramlal College during 1999 is as follows.

Table 6: Percentage of Male and Female Students of BA Class of Ramlal College in 1999

Subject	Percentage of Students	
	Male	Female
Economics	80	20
Geography	68	32
History	62	38
Philosophy	56	44
Psychology	51	49
Sociology	45	55

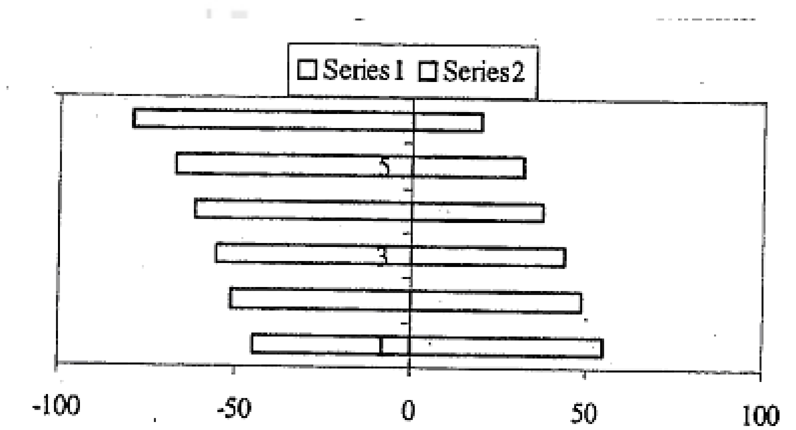


Fig. 6: Male Female Ratio of Arts Students of Ramlal College in 1999

Fig. 6 presents the sliding ratio of the male female students of Ramlal College in 1999

Pyramid Diagram: The uni-dimensional Pyramid Diagram is usually used to show the distribution of male and female by age group. The male and female population of USA in 1950 were as follows:

Table 7: Population of USA by Sex and Age Group in 1950

Age Group	Female (in millions)	Male (in millions)
75 and above	2.1	1.7
70 -74	1.8	1.6
65-69	2.7	2.4
60-64	3.0	3.0
55-59	3.6	3.6
50-54	4.2	4.0
45-49	4.5	4.5
40-44	5.0	5.0
35-39	5.7	5.5
30-34	5.9	5.7
25-29	6.2	6.0
20-24	5.9	5.5
15-19	5.4	5.3
10;14	5.7	5.7
-9	6.4	6.8
Below 5	8.0	8.3

Figure 7 presents the data in pyramidal form. In informetric studies this form of presentation can be employed while studying a particular population, say, scientists, by sex and age group.

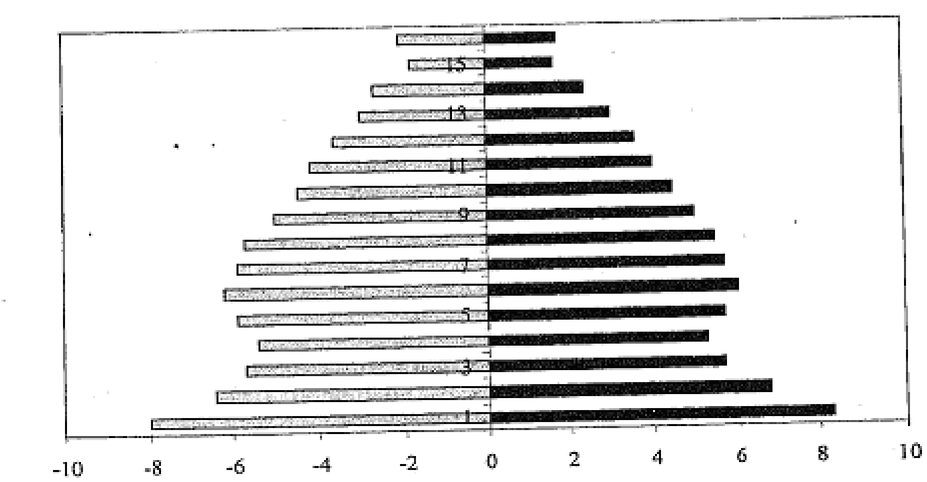


Fig 7: Population of USA in 1950

Two-dimensional Diagrams

These diagrams are also known as area diagrams, as we take both the dimensions, the length and the breadth, that is, the area into account.

1) ***Rectangle:*** Suppose, we intend to represent by rectangles the BA, BCom, and BSc students of Ramlal College of the year 1999 with rectangles. We can do it the following way. The number of students in these three classes respectively are: 1463, 1148, and 669. If 1" in our scale represents 1000, then the rectangles with the sides as 1" x 1.46", 1" x 1.15", and 1" x 0.67" will represent the figures of the three classes. The rectangles in Fig. 8 approximately represent the number of students in each class.

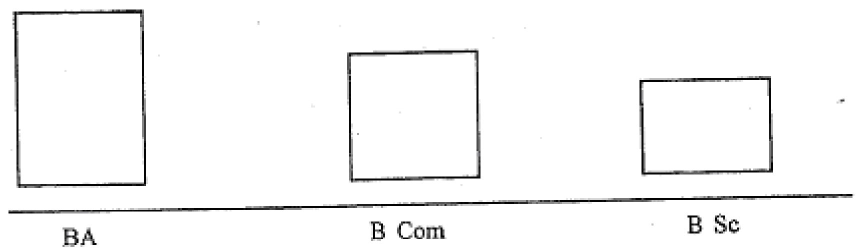


Fig. 8: Undergraduate Students of Ramlal College , 1999

2) ***Square:*** The rectangles in Fig. 8 can also be shown in the form of squares. For drawing the squares, the square roots of all the numbers are computed. In our case, the square roots of the three numbers, 1463, 1148, and 669 are respectively 38.2, 33.9, and 25.9. If we select such a scale where 1" represents 30, then one side of the three squares will respectively be 1.27, 1.13, and 0.86. The squares in Fig. 9, approximately presents the comparative data. Normally, a square diagram may be used when the ratio of the highest value and the lowest value in the set is more than 10:1.

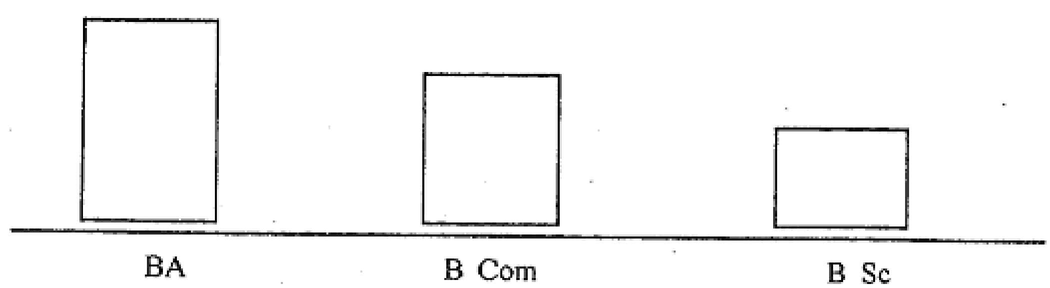


Fig. 9: Undergraduate Students of Ramlal College, 1999

3) ***Circle :*** Now, we shall present the same data in the form of circles in Fig, 10. The area of a circle is given by the formula $7tr^2$. Applying this formula and to the 1463, 1148, and 669, we get the values of the three radii as 21.6, 19.1, and 14.6. Using a suitable scale, we can draw the circles as given in Fig 10.

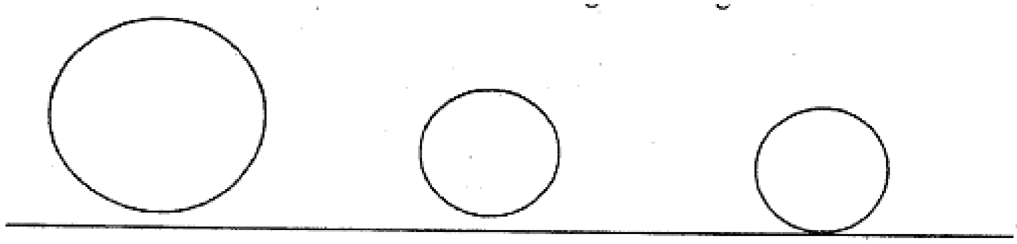


Fig. 10: Undergraduate Students of Ramlal College , 1999

4) ***Pie Chart*** :Here, a circle is divided on the basis of values of angles. The angle value of the subset is given by the following formula:

$$\text{Angle value of students} = \frac{\text{Value of subset}}{\text{Total value}} \times 360$$

The angle values of the number of BA, B Com, and B.Sc. students of Ramlal College of the year 1999 are given by:

$$\text{Angle value of BA students} = \frac{1463}{3270} \times 360 = 161.1$$

$$\text{Angle value of B.Com. students} = \frac{1148}{3270} \times 360 = 126.4$$

$$\text{Angle value of B.Sc. students} = \frac{669}{3270} \times 360 = 73.5$$

With these values, we can draw the circle (pie diagram) as in Fig. 11

1= BA students, 2 = B Corn Students, 3 = B Sc students

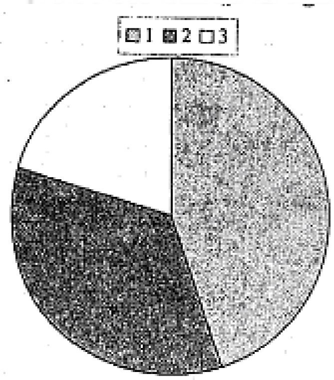


Fig. 11: Undergraduate Students of Ramlal College, 1999

This diagram can also be presented in an exploded form as shown in Fig. 12

1 = BA students, 2= B Com Students, 3 = B Sc students

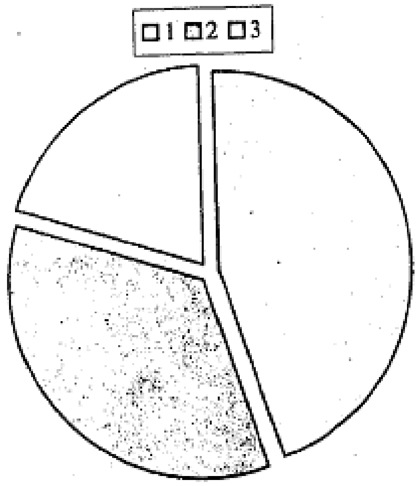


Fig. 12: Undergraduate Students of Ramlal College, 1999

Three Dimensional or Volume Diagrams

Volume diagrams are normally used when the ratio between the highest and lowest values is more than 100 : 1. To draw these diagrams, cube roots of the measurements are computed, and then the sides of each cube is taken in proportion to the cube roots.'The volumes can also be presented in the forms of cuboids, cylinder blocks, pyramids, and so on working out the values using the required mathematical formula. Presenting here are the diagrams 'with the values 1728 and 64.

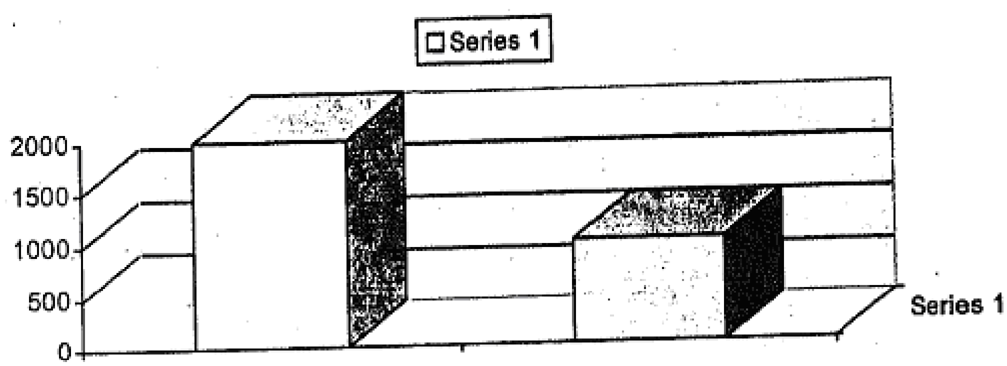


Fig. 13: Presentation in the form of a cube

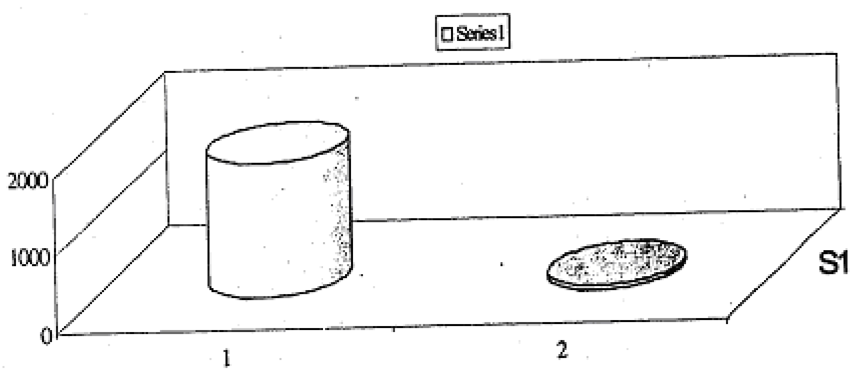


Fig. 14: Presentation in the form of cylinder

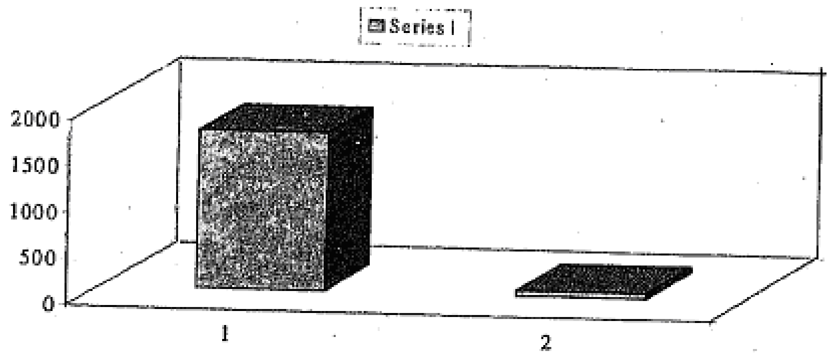


Fig. 15: Presentation in the form of cuboids

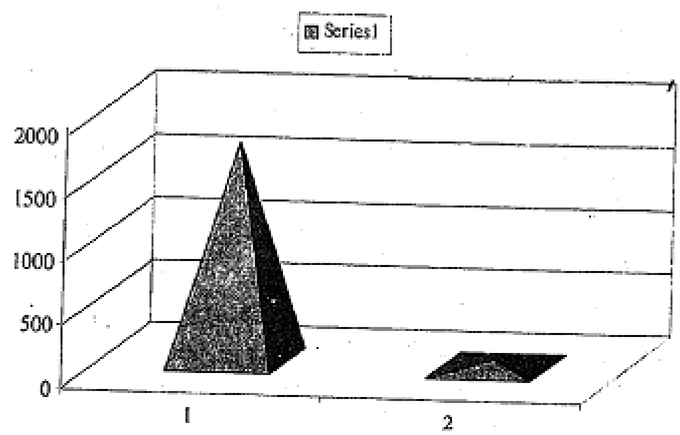


Fig. 16: Presentation in the form of a pyramid

10.8 : Important User Studies

The origin of user studies can be traced back to thel940s. Since then numerous user studies have been conducted all over the world. Sridhar in his *User Research* has provided a chronological bird's eye view of the literature for the period 1940's to 1980's. Devarajan in his *Library Information User and User Studies* has also provided a survey and analysis of user studies reported in *LISA* during 1969 to 1989.

Here, because of the limitation of length of the course material we are just describing briefly three surveys.

One of the first user studies in the field of scientometrics was by Herman H Fussier in 1949. His study appeared in two installments in *Library Quarterly of* the same year. The study comprised two parts - the first part discussed `the importance of the literature *of* various subject fields to research in chemistry and physics' and the second ` the temporal span of literature, the principal forms of the literature, the national origins of the literature used in the United States", and some of the more important serial titles. He advanced the idea of `key' journals and concluded his investigation with the remark: `the distribution of literature as revealed in the study, indicates that provision of a substantial proportion of the references in a research field is possible through a small number of journals if they are properly selected but that the provision of the entire necessary research literature for a field will require a very large number of titles, many of which will be needed elsewhere in a university or in a large research organization'

One of the more important user studies in the field of scientometrics was conducted by the Advisory Council of Scientific Policy, U K in the1960s.The survey was initiated by B H Flowers an,d conducted through questionnaire mailed to as many as 6,195 scientists comprising physicists and chemists in the ratio 1:2. The sample was drawn from membership lists *of* UK-based Chemical Society, Institute of Physics, and Physical Society. The survey yielded around 3,000 responses. The major findings of the survey were as follows:

i) Pure scientists are more dependent on literature compared to industrial scientists and

technologists.

ii) Abstracts are widely used both for keeping uptodate and obtaining specific information.

iii) Patents are not used much by the physicists and chemists. As such there is not much justification of their inclusion in the abstracting journals devoted to pure sciences.

iv) Scientists expressed strong desire for more reviews and for a system of grading references in the original papers according to their utility.

v) The libraries seem to be playing, by and large, a little more than a passive role in the research process, although they are fairly well regarded as a general source *of* information.

A very important study was conducted by the University. of Bath during 1967 to 1970 with a large number of social scientists as the sample. The still-71y is popularly known as INFROSS (Information

Requirements of the Social Scientists). The objective of the study was to gain an overall insight into the varied information requirements of social scientists. Four specific groups studies comprised: i) researchers in social sciences; ii) social scientists in government departments; iii) college of education lecturers and school teachers; and iv) social workers, The project was supported by the Office for Scientific and Technical Information (later the British Library Research and Development Department). The results of the study were published in the form of reports. This project was succeeded by DISISS (Design of Information Services in the Social Sciences) and carried out between 1971 and 1975 with funding from the British Library Research and Development Department.

. 10.9 : Application of User Studies

The results obtained through user studies can be applied in a number of activities. Some of the activities are as follows:

i) Collection development

Through user studies we come to know the actual requirements of the users in terms of books, periodicals, non-book materials, and so on. Knowing this, we can take steps to fill in the gaps, strengthen the collection which is weak, and develop the collection in such-a way that it becomes a really useful collection for the users.

We can stop subscribing to periodicals that are not used or little used. We can discard the items that are no more useful to the users. We can also have a check on the acquisition of non-book materials.

ii) Improvement in the services.

A library provides a number of services to its readers. In case some services are found to be unsatisfactory, measures can be taken to improve the services.

iii) Introduction of new services

If users demand for new services, the matter may be brought to the notice of the authorities, for introduction of those services,

iv) Design of information systems and services

Once the varied requirements of the users and their problems in the use of services are identified, steps

may be taken to design the information services and systems so as to meet the requirements and overcome the problems.

10.10 : Summary

In this Unit, we have discussed the definition and purpose of user studies. The section on methodology deals with various methods: questionnaire, interview and observation, and library record analysis. The section on planning a survey includes the basic objective and scope of the survey, methodology for data collection, framing of the questionnaire along with an example, steps in testing' of the questionnaire, selection of the sample, distribution and receipt of the questionnaire. The classification and tabulation of data, wherein the salient aspects of classification, and the details of the preparation of a table, starting from simple to complex, are described. The diverse ways of data analysis and the presentation of results in tabular and graphical forms are discussed together with a number of examples for uni-dimensional, two-dimensional, and three dimensional graphical presentation. Examples of some important user studies and the application of user studies.highlighted in the final two sections.