


## Subject: Library and Information Science

Production of Courseware

 **-Content for Post Graduate Courses**



**Paper No : 06** Management of Libraries and Information  
Centres and Knowledge Centres

**Module : 28** Statistics and reporting



### Development Team

Principal Investigator  
&  
Subject Coordinator

Dr. Jagdish Arora, Director  
INFLIBNET Centre, Gandhinagar

Paper Coordinator

Prof Dinesh K Gupta  
Professor, V M Open University

Content Writer

Mr P K Bhattacharya  
Fellow, TERI

Content Reviewer

Prof Dinesh K Gupta  
Associate Professor, V M Open University, Kota

## **MODULE 28**

### **STATISTICS AND REPORTING**

#### **I. Objectives**

The objectives of the unit/module are to:

- Explain the concept of statistics;
- Clarify the need of statistics and reporting in libraries;
- Identify methods of data collection for library statistics;
- Discuss about the data representation and reporting format and apply statistics in libraries.

#### **II. Learning Outcomes**

After going this unit/module, you would learn about the need of library statistics and its reporting; concepts of statistics. You would also learn about data collection for library statistics; data representation/reporting format; application of statistics in library, etc.

#### **III. Structure**

1. Introduction
2. Library Statistics and Reporting
  - 2.1 Need of library statistics and its reporting
  - 2.2 Sources of Library Statistics
  - 2.3 Quality of library statistics
  - 2.4 Types of library statistics
3. Concepts of statistics
  - 3.1 What is statistics?
  - 3.2 Why use statistics?
  - 3.3 Statistical terms and their understanding
  - 3.4 Central tendency & deviation
  - 3.5 Measures of Spread
4. Data Collection for Library Statistics
  - 4.1 Data collection methods
  - 4.2 Data quality: LibQual
  - 4.3 Data collection formats
5. Data Representation/Reporting Format
6. Application of Statistics in Library
7. Summary

## 8. Reference

### 1. Introduction

Statistics is a field of mathematics that pertains to data analysis. Statistical methods and equations can be applied to a data set in order to analyze and interpret results, explain variations in the data, or predict future data. A few examples of statistical information we can calculate are:

- Average value (mean)
- Most frequently occurring value (mode)
- On average, how much each measurement deviates from the mean (standard deviation of the mean)
- Span of values over which your data set occurs (range), and
- Midpoint between the lowest and highest value of the set (median)

Statistical methods can be used to determine how reliable and reproducible the measurements are, how much the data varies within the data set, what future projection may be, etc. This module will cover the basic statistical functions of mean, median, mode, standard deviation of the mean, weighted averages which are frequently used in library environment.

Librarians know the value of library as community services, and patrons appreciate their importance as well. But in an increasingly digital world, the role of libraries as community and cultural centers at times are undervalued, and occasionally comes under fire.

Libraries are service based organization. Library functions are carried out on allocated budgets where funds are provided by the governments, funding bodies, donations made by patrons, fees received from members, etc. In strict economic terms, when civil society participants invest money in certain activities it expects some benefits. The case is true for libraries as well. Every library returns back value to the community through its readers and users. Libraries provide both direct and indirect benefits to the society. While the users are directly benefited, however, people associated with the users get indirect benefits from the knowledge, which the library users have gained. In order to make an informed decision regarding very existence of libraries, policy makers need to measure value of libraries. Statistics plays important role in any library to measures extent of benefits delivered to users. While it is difficult to measure indirect values of any library, direct benefits are quite measurable. When values of any library are measured, a cost-benefit analysis can be

done to understand importance of the library, which in other words poses a strong ground for library budget.

In this module, we will focus on how library values can be measured, what are the different tools available for such measurement and what are the best way to present the value measurement report before the competent authority.

## **2. Library Statistics and Reporting**

Statistics refers to a method of dealing with quantitative information involving collection, analysis, presentation and interpretation of data. It is the facts and figures which are presented in tabular or other forms and regarded as one of the important tools for making decisions.

Library statistics are quantitative and qualitative data about library services, library use and library users which are essential for revealing and confirming the outstanding value that libraries provide.

### **2.1 *Need of library statistics and its reporting***

Statistical measures of a library and the services it delivers to users, provide multiple benefits. A policy decision in library can only be taken based on statistical data. Following are some important benefits of library statistics:

- Library statistics are necessary for the effective management of libraries, but they are still more important for promoting library services to the different types of stakeholders including policy makers, funding bodies, library staff, and users.
- Library statistics and subsequent reports are aimed at policy makers, managers and funders for decisions on levels of service and future strategic planning.
- Library statistics can reveal a wealth of material, of hidden success stories where libraries have opened and ensured access to relevant information for all groups of the population.
- By measuring the input into libraries (resources including buildings and equipment, staff and collections), library statistics show the engagement of politics and authorities for library services.
- By counting the output, the usage of traditional and new electronic library collections and services, libraries show that their services are adequate to the respective population. Comparing input and output data demonstrates whether libraries are organising their services in a cost-effective way.

- Data about the use and acceptance of library services can also indicate the outcome of libraries on the user population in terms of literacy, information seeking skills, educational success etc.

Libraries have assumed new responsibilities in a changing information world; they need new statistics for managing and promoting these new tasks. Keeping this in mind library statistics also play pivotal role in following:

- It helps to determine the growth of library as well as planning and controlling the activities of library.
- It helps the librarian for comparison between previous and current library activities.
- It helps librarians for evaluation of the staff performance.
- It helps to write the functions and performance reports of library.
- It also helps in comparing the services offered by a library with other libraries.

## **2.2 Sources of library statistics**

Almost every section of a library has scope of generating statistical data. In order to improve performance, library should set up mechanism to collect data from all possible sources using various survey tools and techniques. For example,

- In acquisition section, periodic records can be kept for how many documents are purchased, collected and accessioned monthwise. Further, processing section can maintain number on books processed, barcoded, etc. These statistics can be used as a measure of staff efficiency.
- In circulation section, monthly records of issuance of books, periodicals can be a measure of library usage.
- In periodical section, a measure of monthwise use of periodicals in print and electronic form could be used for taking a future decision.
- In reference section, users' satisfaction rate could provide a useful measure of library's overall effectiveness and can serve as a guide for service improvement.

Besides, library data can also be collected from other primary and secondary sources:

- Diaries
- Daily, weekly, monthly and quarterly reports
- Gate register
- Written documents of library
- Library software

## **2.3 Quality of library statistics**

While maintaining of quality level in data collection in library is of utmost importance for maintaining quality service and retaining successful customers in a library, it requires a predefined statistical calculation and process to be followed at regular intervals. Correct, reliable and comparable data are crucial for the value and usefulness of library statistics. Following are some of the characteristics that ensure quality in library statistics:

- The quality of library statistics depends on accurate and timely data.
- Data should be collected at macro and micro levels to achieve best understanding and subsequent resolution of problems.
- In order to get accurate data, predefined questionnaires need to be prepared and populated among the target population at periodic intervals.
- Larger the sample population, more accurate will be the results.
- Grouping, formatting and representation of data also play crucial part in reporting library statistics.

#### **2.4 Types of library statistics**

Library statistics are primarily divided into two basic categories – use and user studies. In the first category, use of library resources is studied and in second category, i.e., user studies, library staffs interact with users using different communication modes to understand their feedback.

##### ***Use and user studies in library***

<b>Use studies</b>	<b>User studies</b>
Use studies are conducted to assess the usage pattern of given information sources, such as books, CDs, AV records, databases, and periodical publications. For example. How many times a database is used, how many books are issued, how many CDs are being accessed or how many times the latest issues of particular journals are being read by users, etc. are some questions usually asked by the library to understand the usage pattern of resources it keeps.	A library user study may be defined as any study relating to library use, in any or all of its aspects. These studies aimed at determining the overall pattern of interaction with the user community, without reference to any particular mode of information reception by users.  For example, How many users have visited the library in last month, what is the most important service of the library, how the e-services of library can be improved, how much satisfied you are using your library services, etc. are some of the feedback

	based questions asked from users as part of the user studies.
In general, such data are collected by studying the library records, e.g., books and periodicals used inside or issued from the library.	In this format, periodic feedback of users is taken through some survey tools such as questionnaire, observation or interview.
<p>Use studies are of utmost importance to libraries as they provide:</p> <ul style="list-style-type: none"> <li>• Useful tool to improve operational efficiency.</li> <li>• Most used resources can get priority in collection development process.</li> <li>• Allocation of funding can be reaffirmed based on resource use data.</li> </ul>	<p>User studies are of great importance as they provide:</p> <ul style="list-style-type: none"> <li>• A substantial body of specific knowledge, facts and conclusions.</li> <li>• A user study yields conclusions that can be used in improving the planning and administrative process.</li> <li>• User studies show the different channels employed by users in the information acquisition process and also the different types of information sources and the frequency with which they are used.</li> <li>• User studies indicate whole range of factors that help to determine the nature of the individual information collection process.</li> </ul>

A combination of statistics from both use and user studies help librarians, library committee and library authorities to take decisions regarding assessing performance of library staff, value of library, funding activities under the library, improvement in library services and many more.

### **3. Concepts of Statistics**

#### **3.1 What is statistics?**

Statistics is the methodology for collecting, analyzing, interpreting and drawing conclusions from data and information. Putting it in other words, statistics is the methodology which scientists and mathematicians have developed for interpreting and drawing conclusions from collected data.

Statistics is a discipline that examines data and can calculate numerical estimates of "true" values. Statistics can not prove anything- estimates are normally presented in

probabilistic terms (e.g., we are 95% sure ...), neither statistics can make bad data better as it is like "garbage in, garbage out".

### **3.2 *Why use statistics?***

Statistics can summarize and simplify large amounts of numerical data. By using statistics, one can draw conclusions about data. Statistics may reveal underlying patterns in data not normally observable (especially true in multivariate analyses). If used correctly, statistics can separate the probable from the possible. Suppose one wants to characterize something (species, community composition, stratigraphic range, average grain size, etc...) for which only a limited sample is available - one must estimate the "true" parameters by employing statistical methods. Quantitative and qualitative data about library services, library use and library users are essential for revealing and confirming the outstanding value that libraries provide.

Library statistics are necessary for the effective management of libraries, but they are still more important for promoting library services to the different types of stakeholders, policy makers and funders, library managers and staff, actual and potential users, the media and the general public. Where statistics are aimed at policy makers, managers and funders, they are essential for decisions on levels of service and future strategic planning.

Library statistics can reveal a wealth of material, of hidden success stories where libraries have opened and ensured access to relevant information for all groups of the population.

### **3.3 *Statistical terms and their understanding***

#### ***Population and sample***

Population and sample are two basic concepts of statistics. Population can be characterized as the set of individual persons or objects in which an investigator is primarily interested during his or her research problem. Sometimes wanted measurements for all individuals in the population are obtained, but often only a set of individuals of that population are observed; such a set of individuals constitutes a sample. For example, a research on public library system in India will consider all public libraries in India as population, but some important libraries in Delhi will be called as sample. How big must it be for the sample to represent the population? No real answer as it depends upon the variability of the population and the degree of precision one wants to achieve in answering the question.

<http://www.mv.helsinki.fi/home/jmisotal/BoS.pdf>

***Variables*** A characteristic that varies from one person or thing to another is called a variable, i.e., a variable is any characteristic that varies from one individual member of the



population to another. Examples of variables for humans are height, weight, sex, marital status, and eye color. The first three of these variables yield numerical information (yield numerical measurements) and are examples of quantitative (or numerical) variables, last two yield non-numerical information and are examples of qualitative variables.

Quantitative variables can be classified as either discrete or continuous. A discrete variable is a variable whose possible values are some or all of the ordinary counting numbers like 0, 1, 2, 3, . . . . As a definition, we can say that a variable is discrete if it has only a countable number of distinct possible values. Quantities such as length, weight, or temperature can in principle be measured arbitrarily accurately. However, weight or length may be measured to the nearest gram, but it could be measured more accurately. Such a variable, called continuous, is intrinsically different from a discrete variable.

**Scales** Besides being classified as either qualitative or quantitative, variables can be described according to the scale on which they are defined. The scale of the variable gives certain structure to the variable and also defines the meaning of the variable.

The categories into which a qualitative variable falls may or may not have a natural ordering. If the categories of a qualitative variable are unordered, then the qualitative variable is said to be defined on a nominal scale. On the other hand, if the categories can be put in order, the scale is called an ordinal scale. Based on what scale a qualitative variable is defined, the variable can be called as a nominal variable or an ordinal variable. Examples of ordinal variables are irregular scaled data converted to ranks or relative position, e.g., education (low, high). Nominal or categorical data includes binary data (e.g., presence/absence) or group data (e.g., sandstone/ siltstone/mudstone).

Quantitative variables, whether discrete or continuous, are defined either on an interval scale or on a ratio scale.

Ratio-scale data: Measurements along a continuous scale whose scale begins at 0 (e.g., lengths or widths in mm).

- Interval-scale data: Same as ratio, but data do not have 0 as low end of scale (e.g., temperature).

### ***Organization of the data***

Observing the values of the variables for one or more people or things yield data. Each individual piece of data is called an observation and the collection of all observations for particular variables is called a data set or data matrix. Data sets are the values of variables recorded for a set of sampling units. For ease in manipulating (recording and sorting) the values of the qualitative variable, they are often coded by assigning numbers

to the different categories, and thus converting the categorical data to numerical data. For example, library visitor status might be coded by letting 1,2,3, and 4 denote a person's being daily, weekly, monthly, or occasionally but still coded data still continues to be nominal data.

Data is presented in a matrix form (data matrix). All the values of particular variable is organized to the same column; the values of variable forms the column in a data matrix. Observation, i.e., measurements collected from sampling unit, forms a row in a data matrix.

### **Qualitative variable**

The number of observations that fall into particular class (or category) of the qualitative variable is called the frequency (or count) of that class. A table listing all classes and their frequencies is called a frequency distribution. In addition of the frequencies, we are often interested in the percentage of a class. The percentage can be calculated by dividing the frequency of the class by the total number of observations and multiplying the result by 100. The percentage of the class, expressed as a decimal, is usually referred to as the relative frequency of the class.

Frequency in the class

$$\text{Relative frequency of the class} = \frac{\text{Frequency in the class}}{\text{Total number of observation}} \times 100$$

Total number of observation

A table listing all classes and their relative frequencies is called a relative frequency distribution. One should also state the sample size, which serves as an indicator of the creditability of the relative frequencies. Relative frequencies sum to 1 (100%).

A cumulative frequency (cumulative relative frequency) is obtained by summing the frequencies (relative frequencies) of all classes up to the specific class. In a case of qualitative variables, cumulative frequencies make sense only for ordinal variables, not for nominal variables.

*Example:* Let the library user types of 40 persons are as follows:

R R M W M O R M M M R W R W R R M R R M M M M O M W M M R R M R R M M M R  
M R R O

Summarizing data in a frequency table by using SPSS:

**Table 1:** Frequency distribution of library user types

Users	Statistics	
	Frequency	Percentage
Regular	16	40

Monthly	18	45
Weekly	4	10
Ocassional	2	5
Total	40	100

### ***Quantitative variable***

The data of the quantitative variable can also presented by a frequency distribution. If the discrete variable can obtain only few different values, then the data of the discrete variable can be summarized in a same way as qualitative variables in a frequency table.

If the discrete variable can have a lot of different values or the quantitative variable is the continuous variable, then the data must be grouped into classes (categories) before the table of frequencies can be formed. The main steps in the process of grouping are:

- Find the minimum and the maximum values variable have in the data set
- Choose intervals of equal length that cover the range between the minimum and the maximum without overlapping. These are called class intervals.
- Count the number of observations in the data that belongs to each class ie. class frequency
- Calculate the relative frequencies of each class by dividing the class frequency by the total number of observations in the data.

As a rule of thumb, it is generally satisfactory to group observed values of numerical variable in a data into 5 to 15 class intervals. A smaller number of intervals is used if number of observations is relatively small; if the number of observations is large, the number on intervals may be greater than 15.

Example: Library users age (in years) of 102 people:

34,67,40,72,37,33,42,62,49,32,52,40,31,19,68,55,57,54,37,32,38,20,50,56,48,35,52,29,  
56,68,65,45,44,54,39,29,56,43,42,22,30,26,20,48,29,34,27,40,28,45,21,42,38,29,26,62,3  
5,28,24,44,46,39,29,27,40,22,38,42,39,26,48,39,25,34,56,31,60,32,24,51,69,28,27,38,56,  
36,25,46, 50, 36,58,39,57,55,42,49,38,49,36, 48,44

Summarizing data in a frequency table by using SPSS:

***Table 2: Frequency distribution of library user's age***

Range	Frequency	Percent	Cumulative percent
18-22	6	5.9	5.9
23-27	10	9.8	15.7
28-32	14	13.7	29.4
33-37	11	10.8	40.2
38-42	19	18.6	58.8
43-47	8	7.8	66.7
48-52	12	11.8	78.4
53-57	12	11.8	90.2
58-62	4	3.9	94.1
63-67	2	2.0	96.2
68-72	4	3.9	100.0
Total	102	100.0	

### 3.4 Central tendency & deviation

Plotting data in a frequency distribution shows the general shape of the distribution and gives a general sense of how the numbers are bunched. Several statistics can be used to represent the "center" of the distribution. These statistics are commonly referred to as measures of central tendency.

**Mode** The mode of a distribution is simply defined as the most frequent or common score in the distribution. The mode is the point or value of  $X$  that corresponds to the highest point on the distribution. If the highest frequency is shared by more than one value, the distribution is said to be multimodal. It is not uncommon to see distributions that are bimodal reflecting peaks in scoring at two different points in the distribution.

*Example:* Let us consider the frequency table for library use by 40 library users.

We can see from frequency table that the mode of users type is Monthly as this has the maximum number of user visits.

**Median** The median is the score that divides the distribution into halves; half of the scores are above the median and half are below it when the data are arranged in numerical order. The median is also referred to as the score at the **50<sup>th</sup> percentile** in the distribution. The **median location** of  $N$  numbers can be found by the formula  $(N + 1) / 2$ . When  $N$  is an odd number, the formula yields a integer that represents the value in a

numerically ordered distribution corresponding to the median location. (For example, in the distribution of numbers (3 1 5 4 9 9 8) the median location is  $(7 + 1) / 2 = 4$ . When applied to the ordered distribution (1 3 4 5 8 9 9), the value 5 is the median, three scores are above 5 and three are below 5. If there were only 6 values (1 3 4 5 8 9), the median location is  $(6 + 1) / 2 = 3.5$ . In this case the median is half-way between the 3<sup>rd</sup> and 4<sup>th</sup> scores (4 and 5) or 4.5.

**Mean** The mean is the most common measure of central tendency and the one that can be mathematically manipulated. It is defined as the average of a distribution is equal to the  $\sum X / N$ . Simply, the mean is computed by summing all the scores in the distribution ( $\sum X$ ) and dividing that sum by the total number of scores ( $N$ ). The mean is the balance point in a distribution such that if you subtract each value in the distribution from the mean and sum all of these **deviation scores**, the result will be zero. For example, for the following observations,

Set 1: 1, 4, 4, 5, 7, 7, 8, 8, 9, the mean is 5.89 and the median is 7

Set 2: 1, 4, 4, 5, 7, 7, 8, 8, 542, the mean is 65.11 and the median is 7

This shows that due to the value of observations, median values do not change, however, mean changes.

### 3.5 Measures of Spread

Although the average value in a distribution is informative about how scores are centered in the distribution, the mean, median, and mode lack context for interpreting those statistics. Measures of **variability** provide information about the degree to which individual scores are clustered about or deviate from the average value in a distribution.

**Range** The simplest measure of variability to compute is the range. The range is the difference between the highest and lowest score in a distribution. Although it is easy to compute, it is not often used as the sole measure of variability due to its instability. Because it is based solely on the most extreme scores in the distribution and does not fully reflect the pattern of variation within a distribution, the range is a very limited measure of variability.

**Variance** The variance is a measure based on the deviations of individual scores from the mean. As noted in the definition of the mean, however, simply summing the deviations will result in a value of 0. To get around this problem the variance is based on squared

deviations of scores about the mean. When the deviations are squared, the rank order and relative distance of scores in the distribution is preserved while negative values are eliminated. Then to control for the number of subjects in the distribution, the sum of the squared deviations,  $\sum(X - \bar{X})^2$ , is divided by  $N$  (population) or by  $N - 1$  (sample). The result is the average of the sum of the squared deviations and it is called the variance.

**Standard deviation** The standard deviation ( $s$  or  $\sigma$ ) is defined as the **positive square root** of the variance. The variance is a measure in squared units and has little meaning with respect to the data. Thus, the standard deviation is a measure of variability expressed in the same units as the data. The standard deviation is very much like a mean or an "average" of these deviations. In a normal (symmetric and mound-shaped) distribution, about two-thirds of the scores fall between  $+1$  and  $-1$  standard deviations from the mean and the standard deviation is approximately  $1/4$  of the range in small samples ( $N < 30$ ) and  $1/5$  to  $1/6$  of the range in large samples ( $N > 100$ ).

#### 4. Data Collection for Library Statistics

As discussed earlier, library collects two types of data from use and user studies. While creating a specific format is required for collecting usage data, statistical applications are necessary for carrying out their analysis. Following are the sources from where library usage data are collected:

- Gate register
- Library circulation counter
- Book and periodicals shelving
- Database usage statistics (from in-house and publishers site)
- Accession register
- Financial data eg. Book, periodicals purchase, repairing, furniture acquired etc.
- Electronic services produced

In each of the case, daily, weekly, monthly and annual data are collected and tabulated as per the requirements. The data helps library staffs as well as authority to draw conclusion on usage and performance of the libraries.

While library use statistics requires observation and collection of data in proper format, users' statistics, e.g., Satisfaction rate, usefulness of services, etc. need application of several data collection methods.

#### **4.1 Data collection methods**

Data collection is an essential component for conducting research. Data collection is a complicated and hard task. By and large it is also very difficult to say which is the best method of data collection. While there are different modes and techniques usually employed for collection data, however, some of them are standard in all data collection process and are widely followed. Therefore, which data collection method to use would depend upon the research goals and the advantages and disadvantages of each method.

In order to collect data, the researcher should be able to access the data that needs to be collected for the study. Data can be gathered from a number of sources including written documents, records, workplaces, the Internet, surveys or interviews.

**Observation method** Observation is way of gathering data by watching behavior, events, or noting physical characteristics in their natural setting. Observations can be overt (everyone knows they are being observed) or covert (no one knows they are being observed and the observer is concealed). Observations can also be either direct or indirect. Direct observation is when you watch interactions, processes, or behaviors as they occur; for example, observing a user behaviour when looking for information. Indirect observations are when one watches the results of interactions, processes, or behaviors; for example, measuring the number of books referred by the users at the end of the day.

Observation method is being followed during the following conditions:

- When you are trying to understand an ongoing process or situation.
- When you are gathering data on individual behaviors or interactions between people.
- When you need to know about a physical setting
- When data collection from individuals is not a realistic option.

In order to carry out data collection through observation method, following steps are important:

- Think about the evaluation question(s) one wants to receive answers through observation and select a few areas of focus for your data collection.

- Design a system for data collection. Recording sheets and checklists are the most standardized way of collecting observation data which includes both preset questions and responses.
- Observation guides list the interactions, processes, or behaviors to be observed with space to record open-ended narrative data.
- Field notes are the least standardized way of collecting observation data and do not include preset questions or responses.
- Select an adequate number of sites to help ensure they are representative of the larger population
- Select the observers one may want to include in conducting observations.
- Training the observers is critical for getting relevant data.
- Appropriate time of observations is vital for data collection.

***Interview method*** Interviews are a systematic way of talking and listening to people and are another way to collect data from individuals through conversations. The researcher or the interviewer often uses open questions. Data is collected from the interviewee. The researcher needs to remember the interviewer's views about the topic is not of importance. The interviewee or respondent is the primary data for the study.

Following are some characteristics feature of interview process:

- Good approach to gather in-depth attitudes, beliefs, and anecdotal data from individual patrons.
- Personal contact with participants might elicit richer and more detailed responses.
- Provides an excellent opportunity to probe and explore questions.
- Participants do not need to be able to read and write to respond.
- Requires staff time and quiet area to conduct interviews.
- Requires special equipment.

It is necessary for the researcher to prepare before the actual interview. The interview starts before the interview actually begins. This is the researcher's preparation stage. Once the interview is conducted the researcher needs to make sure that the respondents have:



- A clear idea of why they have been asked for the interview.
- Basic information about the purpose of the interview and the research project of which it is a part.
- Some idea of the probable length of the interview and that one would like to record it (also explaining why).
- A clear idea of precisely where and when the interview will take place.

Also the interview needs to be effective and this is the responsibility of the researcher. The researcher ought to have the following skills and abilities:

An ability to listen

- An ability to be non-judgmental
- A good memory
- Ability to think on his/her feet

Interviews are conducted through direct interaction or by telephonic method. There are many types of interviews, which include:

- Structured interviews, where pre-defined set of same questions are asked to all respondents.
- Semi-structured interviews, these interviews are non-standardized and are frequently used in qualitative analysis. The researcher has a list of key themes, issues, and questions to be covered. In this type of interview, the order of the questions can be changed depending on the direction of the interview.
- Unstructured interviews, are non-directed and is a flexible method. It is more casual than the aforementioned interviews. There is no need to follow a detailed interview guide. Each interview is different. Interviewees are encouraged to speak openly, frankly and give as much detail as possible.
- Non-directive interview. The structured and semi-structured interviews are somewhat controlled by the researcher who has set the issues and questions. In non-directive interviews there is no preset topic to pursue. Questions are usually not pre-planned. The interviewer listens and does not take the lead. The interviewer follows what the interviewee has to say. The interviewee leads the conversation.

**Questionnaire method** Questionnaires can be used to collect data about phenomena that is not directly observable (e.g., inner experiences, opinions, values, interests, etc.). They are more convenient to use than direct observation when used for collecting data on observable behavior.

Data collection through questionnaire has both advantages and disadvantages. The advantages of using questionnaires are as follows:

- Can be given to large groups.
- Respondents can complete the questionnaire at their own convenience, answer questions out of order, skip questions, take several sessions to answer the questions, and write in comments.
- The cost and time involved in using questionnaires is less than with interviews.
- Can include both close-ended and open-ended questions.
- Can be administered in written form or online.
- Personal contact with the participants is not required.
- Staff and facility requirements are minimal.

The disadvantages include:

Inability to probe deeply into respondents' beliefs, attitudes and inner experiences.

- Modifications to the questions can not be made once the questionnaire has been distributed. Interviews typically ask oral questions of individuals.
- Responses are limited to the questions included in the survey.
- Participants need to be able to read and write to respond.
- Takes time to pre-test a written survey to make sure that your questions are clearly stated.
- Relies on participants' perceptions. Be aware of potential gaps between participants' responses and reality.
- Questions on surveys can be misunderstood, especially if they are self-administered and/or if participants do not understand the context for the survey questions.

- Survey questions (especially closed-ended questions) can be limited to what the provider thinks may be the range of responses.

Questionnaire based surveys work better after one has determined the range of outcomes that the survey can target. Therefore, surveys may not be the best initial data collection tool.

A good questionnaire must have certain characteristics. A questionnaire must be:

- Specific : The questionnaire should be concerned with specific topics.
- Short: The questionnaire should be short because very lengthy questionnaires often find their way into the wastebasket.
- Simple and clear: The questionnaire should be clear. As far as possible simple words should be used.
- Objective type questions should be asked.
- Presented in a good order.
- Attractive: A questionnaire must be attractive in appearance.
- Questions should be arranged logically.
- As far as possible personal questions should be avoided.
- The questionnaire must be of convenient size and easy to handle.
- The questions must be arranged in a logical order so that a natural and spontaneous reply follows.
- Instructions with regard to the filling up of the form must be given in the questionnaire itself.
- The number of questions should be kept to the minimum. The number of questions should be limited to the object and scope of the investigation.

It is important to put lot of preparatory research work before developing and populating a questionnaire. Pilot testing is often recommended before wider dissemination of questions. Following are important steps to be followed for developing a good questionnaire and also receiving data successfully:

- Define research objectives: Start with a broad topic, then narrow it by asking five questions about: the time frame, geographical location, conducting a broad

descriptive study versus specifying and comparing different subgroups, what aspect of the topic you want to study, and how abstract is your interest.

- Selecting a sample: Identify the target population from which the sample will be selected.
- Designing the questionnaire (*Appendix I*): keep it short; avoid technical terminology; don't use the term questionnaire or checklist on the form; make it attractive; organize the items so they are easy to read and complete; number pages and items; include return address information on the form and include a postage paid addressed envelope; directions should be clear, brief, in bold print; organize questionnaire in a logical sequence; use a transitional sentence to change topics; begin with interesting, non-threatening topics; put threatening or difficult items near the end; don't put important items at the end of a long questionnaire; be brief in stating each item; avoid negatively stated items; when a general question and a related specific question are to be asked together, ask the general one first; avoid biased or leading questions.
- Additional considerations: Use coding to keep respondent's identity anonymous; design items with respondents in mind so terminology is understood; use a closed form with pre-specified (i.e., multiple choice) response choices.
- Use a scale rather than a one-item test when measuring attitudes. The Likert scale uses a five-point scale ranging from strongly disagree to strongly agree.
- Web questionnaires are easy to use, but measures need to be taken to avoid sampling bias and assure anonymity.
- Pilot-testing the questionnaire: Use a sample from your target population to pilot test the questionnaire. Provide space for them to criticize or make suggestions for improving the items. Ask them to state in their own words what they think each item means. Revise and retest the questionnaire.
- Writing a cover letter: Be brief, explaining the purpose of the study, assuring how confidentiality will be maintained. Include flattery and mention professional affiliations. Mention rewards if you plan to provide them and specify a return date. Attend to the appearance of the cover letter.

- Following up with non-respondents: Send a follow-up letter with another copy of the questionnaire. Stress the importance of the study and the contribution the respondent can make.
- Analyzing questionnaire data: Can use qualitative or quantitative measures to analyze data

## **4.2 Data quality: LibQual**

LibQUAL+ is a suite of services that libraries use to solicit, track, understand, and act upon users' opinions of service quality. These services are offered to the library community by the Association of Research Libraries (ARL). The program's centerpiece is a rigorously tested Web-based survey bundled with training that helps libraries assess and improve library services, change organizational culture, and market the library. The goals of LibQUAL+ are to:

- Foster a culture of excellence in providing library service.
- Help libraries better understand user perceptions of library service quality.
- Collect and interpret library user feedback systematically over time.
- Provide libraries with comparable assessment information from peer institutions.
- Identify best practices in library service.
- Enhance library staff members' analytical skills for interpreting and acting on data.

Library administrators have successfully used LibQUAL+ survey data and statistical interpretation to identify best practices, analyze deficits, and effectively allocate resources. Benefits to libraries include:

- Institutional data and reports that enable one to assess whether the library services are meeting user expectations.
- Aggregate data and reports that allows to compare library's performance with that of peer institutions.

## **4.3 Data collection formats**

Following are some sample data collection formats generally used in library environment to collect statistical information. There may be more or fewer number of formats used for

collection of data in any specific library depending on the nature, services offered and the use pattern.

#### **4.3.1 Data collection through library usage study**

In this section we will look into the formats for collection of data for people who visits library, resources of library and their usage, etc. Some of the formats are shown below to understand the nature of data that library usually stores for betterment of its performance. These sample tables are generally prepared based on parameters specified in SPSS statistical software.

**Sample Table 1: Professional configuration of users**

Sr. No.	Users	No. of Users	Percentage of Total Users
1.	Student		
2.	Researcher		
3.	Professional		
4.	Academician		
5.	Others*		

\* Consultant, Library personnel, Trainee, Professional cum Researcher and Librarian.

**Sample Table 2: Frequency of library users visits**

Sr. No.	Period	Frequency	% of Total users
1.	Regularly		
2.	Weekly		
3.	Monthly		
4.	When needed (WN)		
5.	Total		

**Sample Table 3: Users look into the types of information at the library**

Sr. No.	Varieties of Information	No. of Users	% of Total
1.	Articles/papers (Art/P)		
2.	Specific information on subjects (SIOS)		
3.	News Papers (NP)		
4.	Value added information analysis		

	(VAIA)		
5.	Statistical information (SI)		
6.	Study Material (SM)*		
7.	Books		
8.	A-V documents (A-V D)		

**Note:** \*SM includes Study Material (2), Bibliographic database (2), Journal (2), Fiction (2), Xerox (2), Chemical Engineering Related data on Refinery (1) and MISC Information (1).

**Sample Table 4: Types of knowledge resources referred by the users**

Sr. No.	Resource of Knowledge	No. of Users	% of Total
1.	Books		
2.	Journals/Newsletter (Jls/Nls )		
3.	Databases		
4.	Newspapers		
5.	AV materials		
6.	Others*		

**Note:** \* include CDs, pamphlets, reference documents, reports, etc.

**Sample Table 5: Types of database are used by users**

Sr. No.	Types of Database	No of Users	% of Total Users
1.	Science Direct journal database		
2.	JSTOR		
3.	Springer Link journal databases		
4.	Business databases		
5.	NIC databases		
6.	Other Government databases		
7.	IndiaStats.Com database		
8.	CMIE databases		
9.	Scopus		
9.	Infraline databases		
10.	Others*-I (database that is used by double users only)		

11	Others**-II (database that is used by a single user only)		
12.	<b>Total</b>		100.0

**Note:** \* it includes STN databases, Mathscinet, Tax and Gas Scenario in India, Campus of India, SPICE, Searching books, ETU, OGEL, Emerald insight, J-Gate, SD, MD Consult, OVID, Blackwell-Synergy, Pub Med, Financial, SQL Server, Data on Refinery Process. \*\* it includes Dialog databases, ASME, ASCE, Transport Database, BIS, ASTM, Lexis-Nexis, Factiva, Ebsco, Proquest, First Search (OCLC), Thomson Gale, Hein Online, Books in Print.

#### **4.3.2 User data collection through questionnaire**

When library intends to collect data on users, a questionnaire is generally developed considering the steps discussed above and a questionnaire is floated among the users. Questionnaires may be printed and kept in the library or disseminated through mails or e-mails or can be populated online using any of the quick survey template. Generally, such data are collected as per a defined form and using a statistical package such as SPSS data can be tabulated as per the following samples.

**Sample Table 6: Users' satisfaction level with the services**

Sr. No.	Level of Satisfaction	No of Users	% of Total
1.	Mostly		
2.	Fully		
3.	Often Satisfied (OS)		
4.	Not Satisfied (NS)		
5.	Total		100.00

**Sample Table 7: Response against various types of benefits from databases**

Sr. No.	Whether databases help you in finding information?		Responses about types of benefits from the databases					
			Types of Benefits*		Agree	S. A. **	Total	% of G. Total
	Responses	% of Total Users	1.	TSIRP				
			2.	ADC				
1.	Yes		3.	COD				



2.	No		4.	QRC				
3.	Total	100.0	5.	DWANDA				
			6.	UFIDF				
			7.	UDR				

**Notes:** \*1.TSIRP: Time saver for information retrieval process, 2. ADC: Accuracy of data collected, 3. COD: Comprehensiveness of the database 4.QRC: Quality of the records collected, 5.DWANDA: Data which are normally difficult to access, 6. UFIDF: User-friendly interface and display format, &7. UDR: Updated data regularly. \*\* Strongly Agreed.

**Sample Table 8: User's satisfaction (%)**

Sr. No.	Satisfaction Level	No. Us ers	% of Total Users	Satisfaction Level	% of Total Users
1.	100% (A)			75% and above	
2.	75% (B)				
3.	50% (C )			50% and above	
4.	25% (D)				
5.	Can't say (Ng)				
6.	Total		100.0		

**Sample Table 9: Users' perception on library services**

Sr. No.	Parameters of Library Services	Preferred by no. of users	Strongly Preferred by no. of users	Total Users	% of total users
(1)	(2)	(3)	(4)	(3+4)	(6)
1.	DHLDNH <sup>1</sup>				
2.	SLIAEPWA <sup>2</sup>				
3.	TLSHLUB <sup>3</sup>				
4.	GDRINU <sup>4</sup>				
5.	DULH <sup>5</sup>				
6.	ATSUR <sup>6</sup>				
7.	DAPH <sup>7</sup>				

**Note:** 1: Databases have helped libraries to develop new services, 2: Satisfaction level increased after e-products were added, 3: Trained library staffs handles library users better, 4: Growth of digital resources increased number of users, 5: Database usage in library is high, 6: Average time spent by users is reduced, and 7: Databases are priced high.

## **5. Data Representation/Reporting Format**

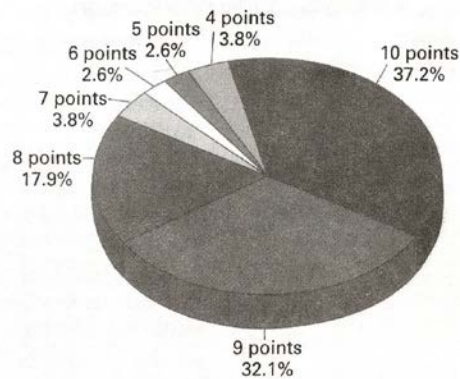
In statistical term, data are either qualitative or quantitative. Depending on the nature of collected data, tables and graphs are drawn as per the following. There are many types of graphs used for representing data, however, we will discuss the most commonly used formats for library data representation.

It is possible to use a mixture of different types of graphs for representing data using statistical packages. However, graphical presentation of data should contain following characteristics:

- Depict the data correctly with minimum distortion.
- Present many numbers in a small place.
- Summarize large data sets and make them coherent.
- Encourage visual comparisons among data elements.
- Have a reasonably clear purpose.
- Make close integration with statistical tables and verbal descriptions.

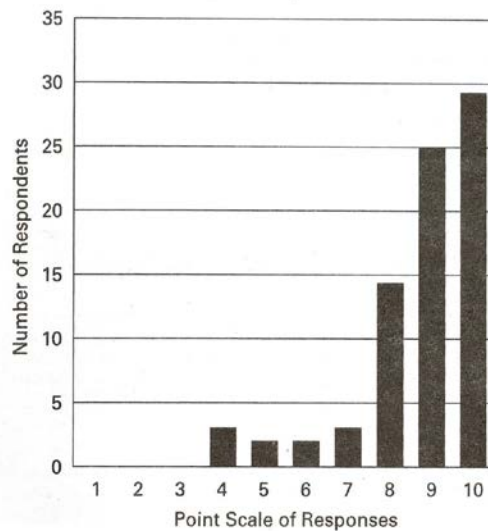
The qualitative data are presented graphically either as a pie chart or as a horizontal or vertical bar graph. A pie chart is a disk divided into pie-shaped pieces proportional to the relative frequencies of the classes. To obtain angle for any class, we multiply the relative frequencies by 360 degrees, which corresponds to the complete circle.

**D. Pie Chart**



On the other hand, a horizontal bar graph displays the classes on the horizontal axis and the frequencies (or relative frequencies) of the classes on the vertical axis. The frequency (or relative frequency) of each class is represented by vertical bar whose height is equal to the frequency (or relative frequency) of the class. In a bar graph, its bars do not touch each other.

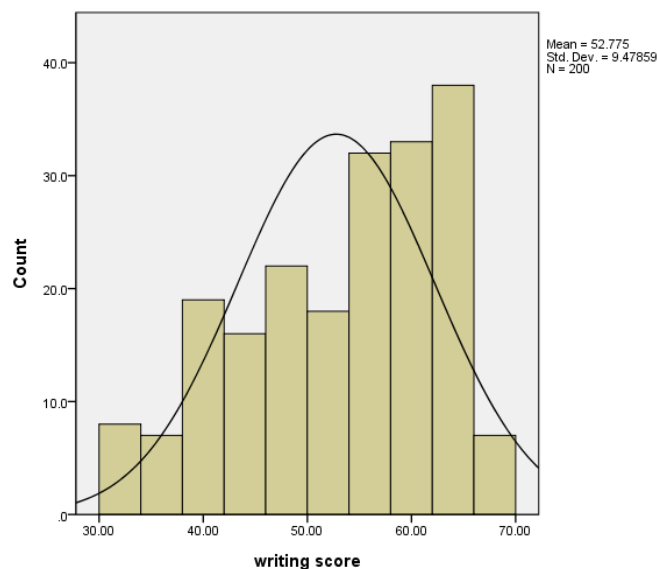
**A. Bar Graph**



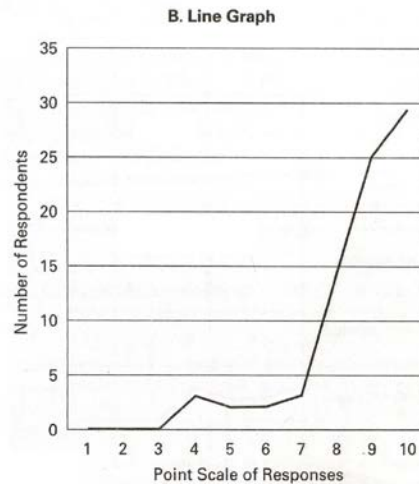
At vertical bar graph, the classes are displayed on the vertical axis and the frequencies of the classes on the horizontal axis. Nominal data is best displayed by pie chart and ordinal data by horizontal or vertical bar graph.

The quantitative data are usually presented in tabular form and then graphically either as a histogram or as a horizontal or vertical bar graph. The histogram is like a horizontal bar graph except that its bars do touch each other. The histogram is formed from grouped data, displaying either frequencies or relative frequencies (percentages) of each class interval. If quantitative data is discrete with only few possible values, then the variable should graphically be presented by a bar graph. Also it is more reasonable to obtain frequency table for quantitative variable with unequal class intervals, then variable should graphically also be presented by a bar graph.

Library data such as monthly statistics of usage of resources, user visits, etc. can also be represented in Histogram. It is a form of a bar graph used with interval or ratio-scaled data. Unlike the bar graph, bars in a histogram touch with the width of the bars defined by the upper and lower limits of the interval. The measurement scale is continuous, so the lower limit of any one interval is also the upper limit of the previous interval.

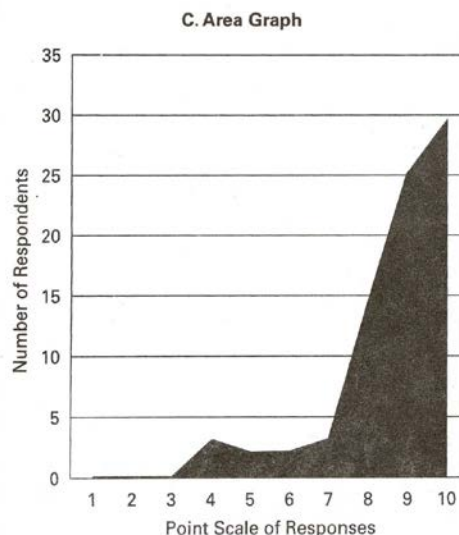


Many a times library data are taken in annualized form, where data are continuous and cumulative. In such cases, Area graphs or Line graphs are used for data reporting.



Lines graphs are similar to frequency polygons, reply of each point and line to epreent different variables of different sets of variables. A line graph an contain more than one line in a graph eg. Displaying types of different resources used in library in a year.

Area graphs are prepared from the line graphs. In this graph, take the line graph and shade the entire area between the axes and the connecting points on the line.

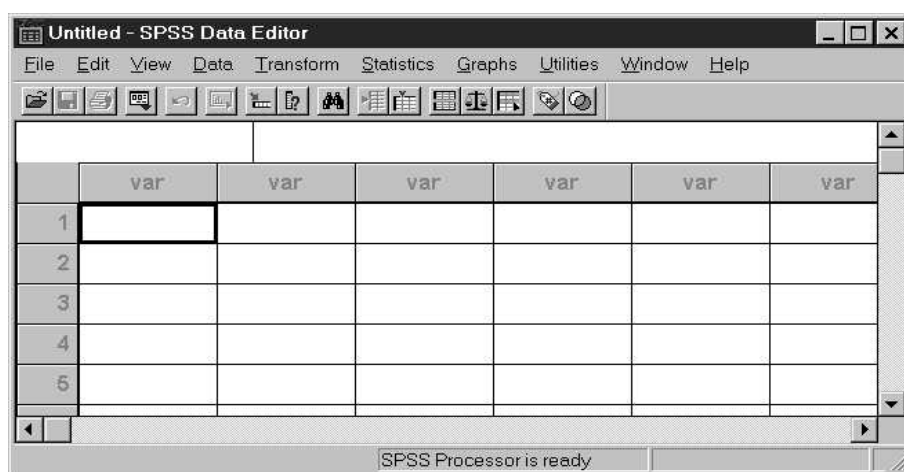


## 6. Application of Statistics in Library

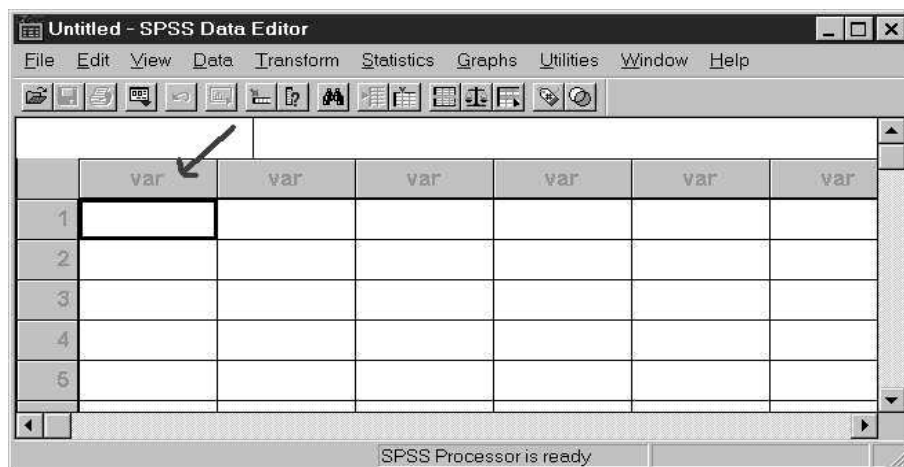
Librarians have been involved in collecting and disseminating statistics for many years. Utilizing statistics to describe and assess the operation of library activities as a part of library tradition has received growing attention from researchers, policymakers, library managers, and professionals.

Once data is collected, data can be entered into a statistical package eg. MS-Excel or SPSS. Following diagram shows the basics of entering data into the SPSS data editor. It has a friendly interface that resembles an Excel spreadsheet and by entering the data directly into SPSS, you don't need to worry about converting the data from some other format into SPSS.

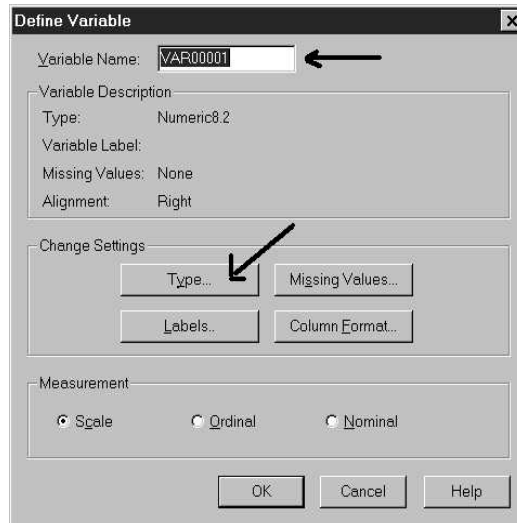
Below is a screen snapshot of what the SPSS data editor looks like when you start SPSS. As you see, it does look like an Excel spreadsheet. In this editor, the columns will represent your variables, and the rows will represent your observations (sometimes called records, subjects or cases).



You can double click on the column heading (shown with the arrow below for the first column). That permits you to enter information about the variable for that column.



The following dialog box allows you to enter information about your variable. For the first variable, let's change the Variable Name to be **Period** (see arrow) and click on **Type** so we can tell SPSS that this is a string variable and click OK.



Once you have created the column headings (variable names) you are ready to enter the data. It is usually best to enter the data one observation at a time going from left to right. After you type in a entry for a variable, you can press the **Tab** key to move to the next variable on the right. Once you reach the last column (**race**) then use the arrow keys to move to the first column of the next observation. Once you have entered the sample data file, the SPSS Data Editor would look like this.

Period	Frequency	Users No.	% of User	price	var
Regularly	53	53	51.5	4099.00	
Weekly	06	06	5.8	4749.00	
Monthly	07	07	6.8	3799.00	
When needed	37	37	35.9	4816.00	
				7827.00	

You can save your data file by clicking **File** then **Save**. It would be wise to save your data about every 10-15 minutes. Imagine spending three hours typing in data, and then the power goes out, your computer stops responding, and then you have to enter the data all over again. The data are not saved when you type them in, rather, the file is saved when you choose **File** then **Save**. The entered data can be retrieved as per the following table

and graph. Librarian can analyse such data for user population in library and their periodicity.

***Sample Table: Frequency of library users' visits***

Sr. No.	Period	Frequency	% of Total users
1.	Regularly	53	51.5
2.	Weekly	6	5.8
3.	Monthly	7	6.8
4.	When needed (WN)	37	35.9
5.	Total	103	100.00



In a questionnaire based data collection approach, a frequency count shows how many people answered a question using a particular response or category. In addition, it is helpful to view the distribution of responses graphically in a bar graph using SPSS software. For example, to examine overall satisfaction, a frequency distribution of the scores of that variable is prepared in the following Table.

In addition to the mean (8.04 in the following table) and the standard deviation for the variable, a horizontal bar graph can provide a visual distribution of scores and also show the mode.

***Frequency of overall user satisfaction scores***

SI No	Count	%	Adj. %	Cumulative %
1	0	0.0	0.0	0.0



2	0	0.0	0.0	0.0
3	1	3.4	3.6	3.6
4	0	0.0	0.0	3.6
5	1	3.4	3.4	7.1
6	2	6.9	7.1	14.3
7	3	10.3	10.7	25.0
8	9	31.0	32.1	57.1
9	8	27.6	28.6	85.7
10	4	13.8	14.3	100.0
Total	28	96.6	100.0	100.0

Mean = 8.04, N of Mean = 28, Std. deviation = 1.59820

The table above shows a fairly high degree of satisfaction. The majority of responses are in the category of 8 and 9 categories, and the category 5 and 6 are below the critical threshold level. The SPSS software allows to filter all responses with an overall satisfaction scores. If the satisfaction survey is repeated next year, it is easy to compare if there is a change in the mean scores. Unlike the median scores mean scores will most likely to reveal changes taken place in library due to actions taken in library for betterment.

Similarly many statistical tables and graphs can be obtained from the data collected using observation, interview or questionnaire methods in library. With the help of MS-Excel, SPSS or LibQual softwares, such data can be used for calculating library performance and guide the authority towards future course of actions.

## 7. Summary

Statistics can be considered as an economic tool that helps libraries to measure the efficiency, performance and service quality and also analyse such data into monetary value of the library for the community. Often such data are presented in report format to the library authority or policy makers or investors to make informed decision regarding future of library in terms of growth, manpower, resources, space, etc. Patrons can evaluate library services based on strong evidence in relation to other available market services, and make decision on existence of the library and the value it offers to the civil society at large.

## 8. References

1. Clayton P and Gorman G E, 2001, Managing Information Resources in Libraries  
London: Library Association Publishing, 272p.
2. Hernon P and Whitman J R. 2009. Delivering satisfaction and service quality: A customer based approach for libraries. New Delhi: Indiana Publishing House, 181p.
3. Durrance J C and Karen E F. 2005., How libraries and librarians help: A guide to identifying user centric outcome. Chicago: American Library Association, 203 p.
4. Jarkko Isotalo, Basics of Statistics  
[www.mv.helsinki.fi/home/jmisotal/BoS.pdf](http://www.mv.helsinki.fi/home/jmisotal/BoS.pdf), accessed on 8 September 2013
5. Research\_Methodology\_Notes  
[http://www.ciilogistics.com/knowledge/Research\\_Methodology/Research\\_Methodology\\_Notes.pdf](http://www.ciilogistics.com/knowledge/Research_Methodology/Research_Methodology_Notes.pdf), accessed on 8 September 2013
6. Barbara M. Wildemuth, 2003. Why Conduct User Studies? The Role of Empirical Evidence in Improving the Practice of Librarianship, INFORUM, 9p.
7. Taylor-Powell E, Steele S, 1996. , Collecting Evaluation Data: Direct Observation.  
<http://learningstore.uwex.edu/pdf/G3658-5.PDF>,  
accessed on 12 September 2013
8. Annabel Bhamani Kajornboon, 2005. , Using interviews as research instruments  
<http://www.culi.chula.ac.th/e-journal/bod/annabel.pdf>, accessed on 8 September 2013
9. SPSS Learning Module: How to input data into the SPSS data editor  
<http://www.ats.ucla.edu/stat/spss/modules/dataed.htm>, accessed on 12 September 2013
10. About LibQual  
<http://www.libqual.org/>, accessed on 12 September 2013
11. IFLA Library Statistics Manifesto  
<http://www.ifla.org/publications/ifla-library-statistics-manifesto>, accessed on 8 September 2013
12. Community Centered: 23 Reasons Why Your Library Is the Most Important Place in Town. <http://publiclibrariesonline.org/>, accessed on 2 December 2013