**UNIT – III**

**CLASSIFICATION OF DATA:** Most research studies result in a large volume of raw data which must be reduced into homogenous groups if we are to get meaningful relationships. This fact necessitates classification of data which happens to be the process of arranging data in groups or classes on the basis of common characteristic. Data having a common characteristic are placed in one class and in this way the entire data get divided into a number of groups or classes.

**TYPES OF CLASSIFICATION:** Classification can be one of the following two types, depending upon the nature of the phenomenon involved:

1. **Classification According to Attributes:** Data are classified on the basis of common characteristics which can either be descriptive or numerical. Descriptive characteristics refer to qualitative phenomenon which cannot be measured quantitatively; only their presence or absence in an individual item can be noticed.
2. **Classification According to Class-intervals:** Unlikedescriptive characteristics, the numerical characteristics refer to quantitative phenomenon, which can be measured through some statistical units. Data relating to income, production, age, weight, etc. come under this category. Such data are known as statistics of variables and are classified on the basis of class intervals.

**BASIC PRINCIPLES IN CLASSIFICATION OF DATA:**

* **Classification should not be Ambiguous:** They very purpose of classification is to remove ambiguity. For this, homogeneity should be made the basis of classification. Various categories / groups should be clearly defined in classification. This avoids confusion in classification and facilitates analysis of concrete nature.
* **Classification should be on the basis of a single classification Principle:** The consideration for classification should be only one at a time. For example, buyers may be classified on the basis of economic or social status or frequency of purchases or the amount spent.
* **Classification should be Mutually Exclusive:** This means the answer given by a respondent should be suitable for placement under one classification only. There should be no overlapping of the categories after classification.
* **Classification should be Mutually Exhaustive:** Classification should cover all data under one category or the other. For this, all possible categories should be prepared at the time of classification. Answer of the respondents should not be omitted in the process of classification.
* **Classification should be Action-Oriented:** Classification should be devised keeping in view the future action to be taken on the basis of the analysis and interpretation of data. Classification is a means and not the end in itself.
* **Classification should be Distinct:** Various sub-divisions under each category of classification should be so distinct as to indicate difference substantial enough to lead to different decisions to suit each specific aspect of the problem.
* **Classification should be pertinent to Marketing Research Project:** The purpose of classification should be to study the marketing problem in the best possible manner. Classification without reference to the purpose will not be useful for analysis of data.

**METHODS OF CLASSIFICATION OF DATA:**

Classification of data is possible on the following basis:

* **Geographical Classification:** In geographical classification, the available data are divided on geographical basis, i.e., area-wise or region-wise. Geographical classification is quite simple, convenient and used extensively.
* **Chronological / Periodical Classification:**  In chronological classification, the data are arranged / classified as per the time or period. Here, year or month is used as a base for classification.
* **Qualitative Classification:** Here, the available data are classified on the basis of some quality, attributes or qualifications.
* **Quantitative Classification:** In quantitative classification, the available data are classified on actual quantitative measurement. This method is also known as “classification by variables”. Variable is a characteristic which changes / varies from observation to observation.

## SOURCES OF DATA

There are two sources of data in [Statistics](https://www.toppr.com/guides/business-mathematics-and-statistics/statistical-description-of-data/introduction-to-statistics/). **Statistical sources**refer to data that are collected for some official purposes and include censuses and officially conducted surveys. **Non-statistical sources** refer to the data that are collected for other administrative purposes or for the private sector.

### ****Collection of Data****

* [Census and Sample Surveys](https://www.toppr.com/guides/economics/collection-of-data/census-and-sample-surveys/)
* [Sampling Errors and NSSO](https://www.toppr.com/guides/economics/collection-of-data/sampling-error-and-the-nsso/)

### Statistical Survey: A statistical Survey is normally conducted using a **sample**. It is also called **Sample Survey**. It is the method of collecting sample data and analyzing it using statistical methods. This is done to make estimations about population characteristics. The advantage is that it gives you full control over the data. You can ask questions suited to the study you are carrying out. But, the disadvantage is that there is a chance of sample error creeping up. This is because a sample is chosen and the entire population is not studied. Leaving out some units of the population while choosing the sample causes this error to arise.

### Census: Opposite to a sample survey, a census is based on ****all items of the population**** and then data are analyzed. Data collection happens for a specific reference period. For example, the [Census](https://www.toppr.com/guides/business-economics-cs/descriptive-statistics/census/) of India is conducted every 10 years. Other censuses are conducted roughly every 5-10 years. Data is collected using questionnaires that may be mailed to the respondents.

Responses can also be collected over other modes of [communication](https://www.toppr.com/guides/business-studies/directing/communication/) like the telephone. An advantage is that even the most remote of the units of the population get included in the census method. The major disadvantage lies in the high cost of data collection and that it is a time-consuming process.

### Register: Registers are basically storehouses of statistical information from which data can be collected and analysis can be made. Registers tend to be detailed and extensive. It is beneficial to use data from here as it is reliable. Two or more registers can be linked together based on common information for even more relevant data collection.

From agriculture to [business](https://www.toppr.com/guides/business-studies/nature-and-purpose-of-business/concept-and-characteristics-of-business/), all industries maintain registers for record-keeping. Some administrative registers also serve the purpose of acting as a repository of data for other statistical bodies in a country.

**PRIMARY DATA**

Primary data is the one, which is collected by the investigator himself for the purpose of a specific inquiry or study. Such data is original in character and is generated by survey conducted by individuals or research institution or any organization.

The objectives of primary data are formulated on the basis of research objectives. Objective set the guidelines and directions of research planning. Formulating the objective offers the best feasible means of solution. The research study should yield measurements related to the research objectives as the measurements will provide directions for a decision.

The findings of the research should be capable of being utilized for the better performance of the organization. The cost-benefits analysis should be made for determining the objectives of the primary data collection.

**METHODS OF PRIMARY DATA COLLECTION**

We collect primary data during the course of doing experiments in an experimental research but in case we do research of the descriptive type and perform surveys, whether sample surveys or census surveys, we can obtain primary data either through observation or through direct communication with respondents in one form or another or through personal interviews.

**SEVERAL METHODS OF COLLECTING PRIMARY DATA:**

* + - 1. Observation Method.
      2. Interview Method.
      3. Questionnaire method and
      4. Schedules.
      5. Other Methods.
* Warranty card
* Distributor audits.
* Pantry audits
* Panels/consumer panels
* Mechanical devices
* Projective techniques
* Depth interviews, and
* Content analysis.

**1. Observation Method:** Observation is an activity of a person which senses and assimilates the knowledge of the phenomenon or the recording of data using instrument. It can also be referred as datum collected during this activity. The observation method is the most commonly used method especially in studies relating to behavioral sciences. In a way we all observe things around us, but this sort of observation is not scientific observation. Observation becomes a scientific tool and the method of data collection for the researchers, when it serves a formulated research purpose, is systematically planned and recorded and is subjected to checks and controls on validity and reliability.

**Observation can be:**

* **Participant Observation Method:** In this method the researcher joints in the daily life of informants or organizations and observes how they behave.

Non-Participant Observation Method: In this method the researcher will not join the informants or organizations but will watch from outside.

1. **Interview Method:** Interview is a conversation between two or more people where question are asked by the interviewer to obtain information from the interviewee. The interview method of collecting data involves presentation of oral-verbal stimuli and reply in terms of oral-verbal responses. This method can be used through personal interviews and, if possible, through telephone interviews.

Interviewing is one of the prominent methods of data collection. It may be defined as a two-way systematic conversation between an investigator and an informant, initiated for obtaining information relevant to a specific study. It involves not only conversation, but also leaning from the respondent’s gestures, facial expressions and pauses, and his environment.

**Types of Interview:**

* **Personal Interview:** Personal or face to face interviewing is a core function of marketing research; much of the quality of the entire research process rests on its effectiveness. Despite the growth in popularity of telephone and mail surveys, personal interviewing retains its long held dominance across a wide spectrum of surveys – market, social, political. A personal interview is face to face communication with the respondent. The interviewer gets in touch with the respondent, asks the questions, and records the answers obtained.
* **Telephonic Interview:** The telephone interview is used when the information to be collected is limited. The telephone interview is used in lieu of personal interviews. It is most frequently used when the information has to be collecting quickly and inexpensively. However, it is not as versatile as personal interview as it is difficult to handle over the telephone information needed probes.

**2. Questionnaire Method:** The questionnaire is a list of questions to be asked from the respondents. It also contains a suitable space where the answers can be recorded. The term questionnaire usually refers to a self-administered process where by the respondent himself reads the question and records his answers without the assistance of an interviewer. This is a narrow definition of a questionnaire.

A questionnaire is a method of obtaining specific information about a defined problem so that the data, after analysis and interpretation, results in a better appreciation of the problem. A questionnaire form, which has to be completed by an interviewer, is often referred as schedule.

* **Decide Questions Sequence:** Once the wording of the individual questions has been determined, it is necessary to set them up in some order. The sequence can influence the results obtained. A questionnaire has three major sections:

i) Basic Information.

ii) Classification Information.

iii) Identification Information.

**3. Schedule:** Schedule is a device in social research, which is most frequently used in collecting field data especially where the survey method is employed. It is used in indirect interview. It contains questions and blank tables, which are to be field in by the investigators themselves after getting information from the respondents. Outwardly schedule and questionnaire appear to be the same but there is some difference between the two. Schedule is used in direct interview and direct observation and is filled in by the research work himself.

According to Goode and Hatt, “Schedule is that name usually applied to a set of questions which are asked and filled in by an interviewer in a face to face situation with another person”.

**5. Other Methods:**

There are a number of methods for collection of data which is described as below:

* **Warranty Cards:** Warranty cards are usually postal sized cards, which are used by dealers of consumer durables to collect information regarding their products. The information sought is printed in the form of questions on the ‘warranty cards’ which is placed inside the package along with the produce with a request to the consumer to fill in the card and post it back to the dealer.

* **Distributor or Store Audits:** Distributor or store audits are performed by distributors as well as manufacturers through their sales men at regular intervals. Distributors get the retail stores audited through salesmen and use such information to estimate market size, market share, and seasonal purchasing pattern and so on. The data are obtained in such audits not by questioning but by observation.
* **Pantry Audits:** Pantry audit technique is sued to estimate consumption of the basket of goods at the consumer level. In this type of audit, the investigator collects an inventory of types, quantities and prices of commodities consumer. Thus in pantry audit data are recorded from the examination of consumer’s pantry.
* **Panels / Consumer Panels:** This form of data collection method is nowadays increasingly used for syndicated research. A panel is a group of study units that exist over time and from which data is collect on a regular interval of time.
* **Use of Mechanical Devices:** The use of mechanical devices has been widely made to collect information by way of indirect means. Eye camera, Pupil metric camera, Psycho-galvanometer, Motion picture camera and Audiometer are the principal devices so far developed and commonly used by modern big business houses, mostly in the developed world for the purpose of collecting the required information.
* **Projective Techniques:** Projective techniques for the collection of data have been developed by psychologist to use projections of respondents for inferring about underlying motives, urges, or intentions which are such that the respondent either resists revealing them or is unable to figure out himself.
* **Depth Interview:** When an interview is held without the aid of a structured questionnaire, the interviewer has freedom in conducting it in the manner he desires. Such interviews are not subject to a well-defined and rigid procedure and are known as informal interviews. They are more appropriate in case of ‘sensitive’ issues, which may require more probing.
* **Content Analysis:** Content analysis consists of analyzing the content of documentary materials such as books, magazines, newspapers and the contents of all other verbal materials, which can be either spoken or printed.

**SECONDARY DATA**

Secondary data are those data which have been already collected and analyzed by some earlier agency for its own use; and later the same data are used by a different agency.

According to W.A. Neiwswanger, “A primary source is a publication in which the data are published by the same authority which gathered and analyzed them.

A secondary source is publication, reporting the data which have been gather by other authorities and for which others are responsible”.

**CHARACTERISTICS FOR SELECTING SECONDARY DATA:**

Secondary data means data that are already available i.e., they refer to the data, which have already been collected and analyzed by someone else. Secondary data may either be published data or unpublished data.

By way of caution, the researcher, before using secondary data, must see that they possess following characteristics:

* **Reliability of Data**: There reliability can be tested by finding out such things about the said data:

1. Who collected the data?
2. What were the sources of data?
3. Were the data collected by using proper methods?
4. At what time were they collected?
5. Was there any bias of the compiler?
6. What level of accuracy was desired? Was it achieved?

* **Suitability of data:** The data that are suitable for one enquiry may not necessarily be found suitable in another enquiry. Hence, if the available data are found to be unsuitable, they should not be used by the researcher. In this context, the researchers must very carefully scrutinize the definition of various terms and units of collection used at the time of collecting the data from the primary source originally.
* **Adequacy of data:** If the level of accuracy achieved in data is found inadequate for the purpose of the present enquiry, they will be considered as inadequate and should not be used by the researcher. The data will also be considered inadequate, if they are related to an area, which may be either narrower or wider than the area of the present enquiry.

**METHODS OF SECONDARY DATA COLLECTION:**

In most of the studies the investigator finds it impracticable to collect first-hand information on all related issues and as such he makes use of the data collected by others. There is a vast amount of published information from which statistical studies may be made and fresh statistics are constantly in a state of production.

**The source of secondary data can broadly be classified under two heads:**

1. Internal Secondary Data.
2. External Secondary Data.
3. **Internal Secondary Data:** Data that originate within the firm for which the research is being conducted are internal data. If they were collected for some other purposes, they are internal secondary data. They may be adapted for the marketing research purposes. They may be formal data and informal data.

Sales analysis and invoicing are considered important source of internal secondary data.

* **Sales Analysis:** Sales analysis is an important tool of marketing research. It is the first step in the marketing research programme and acts as a basis for the development of further marketing research. It reveals the current operating problems in the marketing area where the scope for marketing research can be adequately explored in smaller organizations, sales analysis is an important source of marketing information. It provides a major share of the factors for marketing research.

**Step in Sale Analysis: In sales analysis, the following steps are observed.**

* **Territorial Analysis:** Sales data per territory or region are classified and are put in comparable form to have a bird’s eye view of the total sale. Sales vary from region to region and time to time. The causes of such variations can be revealed by investigation a survey of retailers of the region.
* **Customer Analysis:** Sales data according to different nature of customers are classified and compared. They provide the useful purpose of understanding the nature of customers and their behaviors to the sale.
* **Product Analysis:** Sales data product wise can be significant source of marketing research. The sales may vary from product to product. Understanding of their behavior will be useful for framing product policy.
* **Time Analysis:** Sales data classified as per different segments of time, viz., monthly, six monthly and yearly may be useful source of analysis of sales.
* **Invoice Analysis:** Company invoices have been provided a very useful source of information. A copy of an invoice is preserved and information from it may be punched, tabulated, processed and summarized to provide suitable information to the researcher.
* **Accounting Records:** The basis for accounting records concerned with sales is the sales invoice. The usual sales invoice has a sizable amount of information on it, which generally includes name of customer, location of customer, items ordered, quantities ordered, quantities shipped, dollar extensions, back orders, discounts allowed, and date.

1. **External Secondary Data:**

The second forms of secondary data are external source which are generally published and are available in different forms and form different sources. Although external secondary data may be obtained from different sources, some of the sources are given here.

* **Libraries:** Researchers first attend libraries to find out relevant data pertaining to research. They provide many sources where suitable data may be obtained. Public libraries and college and University libraries contain a large amount of business information, which provides sources of other data. Management books, theses, management journals and other publications can be consulted in these libraries.
* **Literature:** A great amount of secondary data is available from literature, particularly literature on marketing subjects. With the development of marketing researches in different countries, new and interesting facts are coming into picture, which are available in various publications. Consultations of this literature may provide proper guidance pertaining to publication, which can be used from time to time.
* **Periodicals:** Business periodicals published fortnightly, monthly, quarterly, semi annually and annually are often consulted by the marketing executive and researchers to plan and design their marketing research. Also to use the available data for research purposes and to verify the conclusions derived from the marketing research, especially of field research. Periodical economic abstracts on economics, finance, trade, transport, industry, labor and management are being prepared by the Government as well as by the non-government agencies.
* **Census and Registration Data**: Census and Registration data have become very comprehensive sources of marketing research. Previously, these concentrated only one population census, but it now extends too many areas.
  1. Census of Population.
  2. Census of Agriculture.
  3. Census of Cattle.
  4. Census of Trade.
  5. Census of Transport.
  6. Census of Industry.
  7. Census of Banking and Finance.
  8. Registration Data.
* **Trade Association:** Trade Association may be an excellent source of data pertaining to an industry. The trade association of one industry may exchange data with the trade association of another industry, and within one industry a firm may exchange data with another firm with the help of trade association of the industry.
* **Government Departments:** Different Government departments have different data, which are not available in libraries. But these are very useful for understanding various aspects of the economy. The researchers can utilize them for the purposes of their researchers.
* **Private Source:** Private Sources include varied sources available in the form of books, monographs, bulletins, journals, commercial reports and so on. They are priced and publicly circulated. Some of the sources include extensive original research, and some summarize the research findings of other person. Many of them are statements of cats and opinions.
* **Commercial Data:** There are several institutions and companies, which purchase and sell marketing information and data. Some of the companies are solely engaged in marketing research. They collect information and data directly from field surveys. Some such companies collect and process secondary data and supply them to their subscribers.
* **Financial Data:** The financial data of reputed concerns are available in several magazines, newspapers, journals and in summary of statistics. The Directorate of Income Tax publishes information pertaining to taxes and income ranges. Such information and data are useful to forecast the market potential of a particular product.
* **International Organizations:** International Organizations such as the International Monetary Fund, the World Bank, the United Nations Organizations, the Asian Ban, the African Bank, Foreign embassies etc. publish several useful statistics, which can be used by researchers.
* **References and Bibliography:** In every publication, the researcher can find references and a bibliography which can be very good sources of information of marketing research. The researcher can consult them for further information and data.
* **Volumes of Statistics:** There are several private and public organizations, which prepare a summary of statistics. In India, the Indian Statistical Institute publishes the Statistical Abstract. Commerce Pvt. Ltd. The Times of India Ltd. And the Financial Express compiles Directories of different subjects. The Government of Indian publishes the Economic Survey of India wherein statistics relating to every field of economic activities are compiled in a suitable form.
* **Advertising Agencies:** Advertising agencies have proved to be very useful sources of marketing research. Recently, a large number of agencies have come into the findings of the advertising researches for their clients. Advertising agencies sometimes, publish reviews, resumes and tests of marketing researches.
* **Other Sources:** There are several other sources of marketing researches. Individuals conduct their own researches, which may be purchased by other institutions. Marketing Associations, Management Associations and individual business houses have been conducting marketing researches for other researches.

**DATA PROCESSING**

The data, after collection, has to be processed and analyzed in accordance with the outline laid down for the purpose at the time of developing the research plan. This is essential for a scientific study and for ensuring that we have all relevant data for making contemplated comparisons and analysis. Technically speaking, processing implies editing, coding, and tabulation of collected data so that they are amenable to analysis.

The term analysis refers to the computation of certain measures along with searching for patterns of relationship that exist among data-groups. Thus, “In the process of analysis, relationships or differences supporting or conflicting with original or new hypothesis should be subjected to statistical tests of significance to determine with what validity data can be said to indicate any conclusions”.

**Processing Operation:**

Processing of data includes:

* 1. Editing,
  2. Coding,
  3. Classification, and
  4. Summarization of data (Tabulation and graphic presentation)

**EDITING:**

The raw data is likely to contain a number of errors during the process of recording the information in surveys. By means of editing one tries to eliminate the errors or remove the points of confusion, if any.

Alternately, recorded raw data is normally less than perfect and the first phase through which this data must pass is editing.

**ESSENTIALS OF EDITING:**

* **Completeness:** The editor must find whether there is an answer to every question. If the interviewer has forgotten to ask a question or to record the answer; it may be possible to deduce from other data on the questionnaire and fill up the gap. The interviewer may also be able to fill in the gap from his memory. If the information is vital, it can be collected by a postal enquiry.
* **Accuracy:** The Editor must try to check whether the answers are accurate. In the first place, inconsistencies should be looked for. Every effort should be made to resolve clear inconsistencies. If this not possible, then the inconsistent entries should be omitted.

Data Processing

Refined Coding Computer Feeding

Data Distribution

Tabulation

Univariate Bivariate Multivariate

**Data Analysis**

Classification Frequency Measurement

Distribution

Data Interpretation

Diagrammatic Presentation

* **Uniformity**: In editing data sets, another keen look out should be for any lack of uniformity in interpretation of questions and instructions by the data recorders. For instance, the responses towards a specific feeling could have been queried from a positive as well as a negative angle.

**STAGES OF EDITING:**

The editing may be done in two stages:

1. **Field Editing:** The field editingis a preliminary editing done to detect the glaring omissions and inaccuracies in the data. It is useful for controlling the filed force and removing misunderstanding. The field editing is done immediately after collection of data because the interviewers then have a fresh memory about the lapses and wrong statements of answers.
2. **Office Editing:** The office editing is done after the field editing. This implies a complete and through scrutiny of the questionnaire. There should be expert editors in the office to evaluate and examine the completed returns of the respondents. There may be a division of labor for editing purposes, and each editor may be assigned a particular portion of the questionnaire.

**FACTORS TO BE TAKEN INTO CONSIDERATION WHILE EDITING:**

* **Fictitious Interviews:** The interviewer himself may complete the questionnaire without contacting the respondent. Sometimes poor quality of interviewing and poor quality responses are the main causes of errors in raw data, i.e., some interviewers may deliberately fabricate the data and fill these in the questionnaire or other forms without practically interviewing the respondents.
* **Erroneous Information and Inconsistencies:** There may be some answers which may be contradicting or inconsistent with another answer in the same questionnaire e.g. in one part of the questionnaire it may be listed that there are no children in household whereas at some other place while listing the ages of household members there may be ages of some children.
* **Responses Not Legible:** If the interviewer or the respondent’s hand writing is bad then the recorded information may not be legible in some cases. The larger the number of open-end questions, the more prominent will be the problem of stops or illegible responses.
* **“Don’t know” and “No Answer”:** Sometimes the respondent answer to a particular question is “Don’t know”.
* **Poorly Designed Questionnaire:** In such cases there is little remedy for missing questions / answers.
* **Inadequate Answers:** The answers supplied by the respondent are either incomplete or ambiguous.
* **Irrelevant Answers:** Here the answer provided by the respondent is out of context. This may be due to the inability o respondent to understand the question properly or he may not be competent to provide required information properly. Intensive questioning may lead to condition a respondent on some particular subject.

**TYPES OF EDITING:**

* **Manual Editing:** Here the editor scans the whole data and amends it in accordance with the given instructions. Here the drafted instructions are of great significance and helps the editor to perform his work systematically with littler scope of using his own judgment and intuition.
* **Mechanical Editing:** When data are transferred on punch cards then counter sorter can be used for editing the raw data. It provided exhaustive and through logical checking. Cards which contain some particular type of error/errors are sorted out. The process is extremely quick and effective. Due to this it is found to be cheap and free from error.
* **Computer Editing:** It is similar to mechanical editing with the difference that several hundred editing instructions can be executed at one time on a computer. Thus here editing is faster and accurate.

**CODING:** Coding is the procedure of classifying the answers to a question into meaningful categories. The symbols used to indicate these categories are called codes. Coding refers to the process of assigning numerals or other symbols to answers so that responses can be put into a limited number of categories or classes. Such classes should be appropriate to the research problem under consideration. They must also possess the characteristic of exhaustiveness and also that of mutual exclusivity which means that a specific answer can be placed in one and only one cell in a given category set. Another rule to be observed is that of unit-dimensionality by which is meant that every class is defined in terms of only one concept.

**ESSENTIALS OF CODING:**

* **Appropriate to the Research Problem:** The number of categories should be reasonable. Generally 3 to 5 categories are considered reasonable. For example, categorization of respondents on the basis of age, say, under 20 years, 20 to 29 years, 30 to 39 years and above 40 years may be relevant to the study.
* **Exhaustive:** The rule on exhaustiveness states that there must be a category for every item.
* **Mutually Exclusive:** In an occupation survey the classification of respondents may be professional, managerial, sales, clerical, craftsman, operatives and unemployed. Some respondents will think of themselves as being in more than one of these groups.
* **Single Dimension:** Every class in the category set is defined in terms of one concept. Suppose a respondent is both a salesman and unemployed. The ‘Salesman’ label expresses the concept of occupational type while the response unemployed is another dimension.
* **Code Sheets:** Code sheets instruct the coders how to handle each question. When codes are defined, a listing of all types of responses being assigned to each code should be made on the code sheet. The code sheet consists of the following elements:

i) **Heading:** The heading identifies the study.

ii) **Column Identification:** Each applicable column is to be listed on the left hand side of the page.

iii) **Coding Instructions:** Limitations / Restrictions and required multiple codes are to be stated next to each code where applicable.

iv) **Special Instructions:** Instructions for the use of ‘borrowed columns’

**CLASSIFICATION OF DATA:**

Most research studies result in a large volume of raw data which must be reduced into homogenous groups if we are to get meaningful relationships. This fact necessitates classification of data which happens to be the process of arranging data in groups or classes on the basis of common characteristic. Data having a common characteristic are placed in one class and in this way the entire data get divided into a number of groups or classes.

**TYPES OF CLASSIFICATION:**

Classification can be one of the following two types, depending upon the nature of the phenomenon involved:

**Classification According to Attributes:** Data are classified on the basis of common characteristics which can either be descriptive or numerical. Descriptive characteristics refer to qualitative phenomenon which cannot be measured quantitatively; only their presence or absence in an individual item can be noticed.

**Classification According to Class-intervals:** Unlikedescriptive characteristics, the numerical characteristics refer to quantitative phenomenon, which can be measured through some statistical units. Data relating to income, production, age, weight, etc. come under this category. Such data are known as statistics of variables and are classified on the basis of class intervals.

**BASIC PRINCIPLES IN CLASSIFICATION OF DATA:**

* **Classification should not be Ambiguous:** They very purpose of classification is to remove ambiguity. For this, homogeneity should be made the basis of classification. Various categories / groups should be clearly defined in classification. This avoids confusion in classification and facilitates analysis of concrete nature.
* **Classification should be on the basis of a single classification Principle:** The consideration for classification should be only one at a time. For example, buyers may be classified on the basis of economic or social status or frequency of purchases or the amount spent.
* **Classification should be Mutually Exclusive:** This means the answer given by a respondent should be suitable for placement under one classification only. There should be no overlapping of the categories after classification.
* **Classification should be Mutually Exhaustive:** Classification should cover all data under one category or the other. For this, all possible categories should be prepared at the time of classification. Answer of the respondents should not be omitted in the process of classification.
* **Classification should be Action-Oriented:** Classification should be devised keeping in view the future action to be taken on the basis of the analysis and interpretation of data. Classification is a means and not the end in itself.
* **Classification should be Distinct:** Various sub-divisions under each category of classification should be so distinct as to indicate difference substantial enough to lead to different decisions to suit each specific aspect of the problem.
* **Classification should be pertinent to Marketing Research Project:** The purpose of classification should be to study the marketing problem in the best possible manner. Classification without reference to the purpose will not be useful for analysis of data.

**METHODS OF CLASSIFICATION OF DATA:**

Classification of data is possible on the following basis:

* **Geographical Classification:** In geographical classification, the available data are divided on geographical basis, i.e., area-wise or region-wise. Geographical classification is quite simple, convenient and used extensively.
* **Chronological / Periodical Classification:**  In chronological classification, the data are arranged / classified as per the time or period. Here, year or month is used as a base for classification.
* **Qualitative Classification:** Here, the available data are classified on the basis of some quality, attributes or qualifications.
* **Quantitative Classification:** In quantitative classification, the available data are classified on actual quantitative measurement. This method is also known as “classification by variables”. Variable is a characteristic which changes / varies from observation to observation.

**TABULATION AND GRAPHIC PRESENTATION:**

Tabulation is the primary function of Data Analysis. The data is validated and analyzed to generate tables in a client-specified format that helps the researcher to interpret the result of the survey and present it to his / her client.

In a graphic mode of presentation the points or lines of various kinds are used to represent data. Each graph paper has thick lines for each division of an inch or centimeter and thin lines for smaller parts of the same.

**Tabulation:** Tabulation may be defined as the logical and systematic organization of statistical data in rows and columns. It is designed to simplify presentation and facilitate analysis.

According to Tuttle, “The logical listing of related quantitative data in vertical columns and horizontal rows of numbers with sufficient explanatory and qualifying words, phrases and statements in the form of titles, heading and explanatory notes to make clear the full meaning, context and the origin of the data”.

According to Secrist, “Tables are a means of recording in permanent form the analysis that is made through classification and by placing in juxtaposition things that are similar and should be compared.”

**CHARACTERISTICS OF A GOOD TABLE:**

* **Suit the Purpose:** A table should be in keeping with the object of statistical enquiry.
* **Scientifically Prepared:** The table should be prepared in a systematic and logically organized manner, simple and compact so that it is readily comprehensible. It should be free from all sorts of over-lapping and ambiguities.
* **Clarity:** A table should be easily understandable, complete and self/explanatory.
* **Manageable Size:** A table should be so designed that it is neither very long and narrow nor very short and broad. If need be, it should be adjusted to the space provided for the purpose.
* **Columns and Rows should be Numbered:** When there are a number of rows and columns in a table, they must be numbered for reference.
* **Suitably Approximated:** If the figures are large, they should be suitably approximated or rounded. The method of rounding should be indicated along with the units of measurement such as a weight in thousand tones rounded to the nearest whole number.
* **Attractive Get-Up:** A table should have an attractive get-up which is appealing to the eye and the mind so that the reader may grasp it without any strain. The rows and columns are separated by single, double or thick lines depending on the broad classes and sub-classes used.
* **Units:** The unit designations should be given at the top of the table below the title such as ‘price in rupees’ and ‘weight in tones’. If there are different units for different items them they should be mentioned in respective columns a row.
* **Averages and Totals:** The averages and percentages should, as far as possible, be given to the right or at the bottom of the columns containing original figures. Totals and sub-totals of both columns and rows and if necessary, the cross totals of each such group should be given.
* **Logical Arrangement of Items:** There should be logical and systematical classification of items in the table. Items may be arranged.

i) Alphabetically,

ii) Geographically,

iii) Chronologically,

iv) Conventionally,

v) In order of magnitude, in ascending or descending order.

* **Proper Lettering:** It is notadvisable to use too many styles of letters in a table. Large capital letters may be used for prefatory notes, footnotes and, source notes. Lettering also helps in adjusting the size of the table.

**STRUCTURE OF TABLE OR MAIN PARTS OF A TABLE:**

* Table Number: Each table should be numbered. There are different practices with regard to the place where this number is to be given. The number may be given either in the centre at the top above the title or inside of the title at the top or in the bottom of the table on the left-hand side.

* **Title of the Table:** Every table must be given suitable title. The title is a description of the contents of the table. A complete title has to answer the question what, where and when in that sequence.
* **Caption:** Caption refers to the column headings. It explains what the column represents. It may consist of one or more column headings. Under a column heading there may be sub-heads. The caption should be clearly defined and placed at the middle of the column.
* **Footnotes:** Anything in a table, which the reader may find difficult to understand from the title, captions and stubs should be explained in footnotes. If footnotes are need they are placed directly below the body of the table.
* **Stubs:** These refer to the heading of the horizontal rows and they are written on the left hand side of the rows. Whether there is need for stubs, and if yes, how many, world depend on the nature of the data.
* **Body:** The body of the table contains the statistical data, which have to be presented. This is the most vital part of a table and the data contained in the body are arranged according to captions and the stubs.
* **Head note:** Statistical tables contain a head note, which refers to the data contained in the major part of the table, and it is placed below the title of the table. Generally written as head note like ‘in lakhs’ or ‘in tones’ etc.,

**TYPES OF TABLES:**

Table may broadly be classified into tow categories:

1. Simple and Complex tables.
2. General purpose and special purpose tables.
3. **Simple and Complex Tables:** The distinction between simple and complex table is based upon the number of characteristics studied. In a simple table only one characteristic is shown. Hence, this type of table is known as one-way table. In a complex table, on the other hand, two or more characteristics are shown. Such tables are more popular in practice because they enable full information to be incorporated and facilitate a proper consideration of all related facts.

* **Simple Table or One-way Table:** In this type of table only one characteristic is shown. This is the simple type of table.
* **Two-way Table:** Such a table shows two characteristics and is formed when either the stub or the caption is divided into two coordinate parts.
* **Higher Order Table**: When three ore more characteristics are represented in the same table, such a table is called higher order table. The need for such a table arises when we are interested in presenting a number of characteristics simultaneously.

1. **General and Special Purpose Tables:** General purpose tables also known as reference tables or repository tables provide information for general use or reference. They usually contain detailed information and are not constructed for specific discussion. In other words, these tables serve as repository of information and are arranged for easy reference.

**GRAPHIC PRESENTATION:**

A graph is a visual form of presentation of statistical data. A graph is more attractive than a table of figure. Even a common man can understand the message of data from the graph. Comparisons can be made between two or more phenomena very easily with the help of a graph.

In a graphic mode of presentation the points or lines of various kinds are used to represent data. Each graph paper has thick lines for each division of an inch or centimeter and thin lines for smaller parts of the same.

**GENERAL RULES FOR GRAPHIC PRESENTATION:**

While graphing statistical data, the following points should be borne in mind.

* **Title:** Every graph must have a clear and comprehensive title so that what facts are presented in the graph may be known.
* **Structural Framework:** The independent variables should always be measured along the X-axis and dependent variables along the Y-axis. The scale on Y-axis should begin from zero as origin. For actual plotting of the data, it should be remembered that every value of the independent variable, there is a corresponding value of the dependent variable.
* **Choice of Scale:** It should be made such that it accommodates the whole data. According to A.L. Bowley, “It is difficult to lay down rules for proper choice of scales by which the figures should be plotted out. It is only the ration between the horizontal and vertical scale that need to be considered. The figure must be sufficiently small for whole of it to be visible at once; if the figure is complicated or related to long series of years and has varying numbers, minute accuracy can be sacrificed.
* **Use of False Base Line:** When fluctuations in a variable are small relative to its size and it is desired to view these fluctuations properly, the vertical scale may be stretched. This can be done if, instead of showing the entire scale from zero to the highest value, only as much need be shown as is necessary for the purpose. The portion which lies between zero and the lowest value of the variable is left out. This omission is indicated by a scale break.
* **Use of Ratio or Logarithmic Scale:** For showing proportional changes, ratio or logarithmic scale should be sued.
* **Line Design:** If more than one line is plotted on the same graph, it is necessary to distinguish those by different patterns as shown in the figure.

* **Caption:** The scale caption for the X-axis is placed under the centre of the horizontal axis. The scale caption for the Y-axis is placed at the top or the middle of the Y-scale.
* **Index:** An index should be given to show the scale and the meaning of different curves.
* **Source:** The source noted should indicate the source of information.

**TYPES OF GRAPHICAL DEPICTION:**

Some important types of graphs which are more popular and they are:

1. **Histogram:** The word histogram is derived from Greek; histos ‘anything set upright’ (as the masts of a ship, the bar of a loom, or the vertical bars of a histogram); gramma ‘drawing, record, writing’. In statistics, a histogram is a graphical display of tabulated frequencies, shown as bars. It shows what proportion of cases fall into each of several categories.
2. **Frequency Polygon:** If we mark the midpoints of the top horizontal sides of the rectangles in a histogram and join them by a straight line, the figure so formed is called a Frequency Polygon. This is done under the assumption that the frequencies in a class interval are evenly distributed throughout the class.
3. **Ogives or Cumulative Frequency Curves:** When cumulative frequencies are plotted on a graph, then the frequency curve obtained is called ‘ogive’ or ‘cumulative frequency curve’. Ogives determine media, quartiles, percentiles, etc.

**Methods: Construction of Ogives**

* **Less than Ogive:** In less than ogive method we start with the upper limits of the classes and go adding the frequencies. When these frequencies are plotted, we get a rising curve.
* **More than Ogive:** In more than ogive method, we start with the lower limits of the classes and from the total frequencies we subtract the frequency of each class. When these frequencies are plotted we get a declining curve.

1. **Pie Chart:** Another way of preparing a two-dimensional diagram is in the form of circles. A pie chart (or a circle graph) is a circular chart divided into sectors, illustrating relative magnitudes or frequencies or percents. In a pie chart, the are length of each sector (and consequently its central angle and area), is proportional to the quantity it represents.

**FORMULATION OF HYPOTHESIS**

**Introduction:** A hypothesis is used to explain a phenomenon or predict a relationship in communication research. There are four evaluation criteria that a hypothesis must meet. First, it must state an expected relationship between variables. Second, it must be testable and falsifiable; researchers must be able to test whether a hypothesis is truth or false. Third, it should be consistent with the existing body of knowledge.

**FORMULATION AND CHARACTERISTICS OF HYPOTHESIS**

1. Hypothesis should be clear and precise. If the hypothesis is not clear and precise, the inferences drawn on its basis cannot be taken as reliable.
2. Hypothesis should be capable of being tested. In a swamp of untestable hypotheses, many a time the research programmes have bogged down. Some prior study may be done by researcher in order to make hypothesis a testable one. A hypothesis “is testable if other deductions can be made from it which, in turn, can be confirmed or disproved by observation.”
3. Hypothesis should state relationship between variables, if it happens to be a relational hypothesis.
4. Hypothesis should be limited in scope and must be specific. A researcher must remember that narrower hypotheses are generally more testable and he should develop such hypotheses.
5. Hypothesis should be stated as far as possible in most simple terms so that the same is easily understandable by all concerned. But one must remember that simplicity of hypothesis has nothing to do with its significance.
6. Hypothesis should be consistent with most known facts i.e., it must be consistent with a substantial body of established facts. In other words, it should be one which judges accept as being the most likely.
7. Hypothesis should be amenable to testing within a reasonable time. One should not use even an excellent hypothesis, if the same cannot be tested in reasonable time for one cannot spend a life-time collecting data to test it.
8. Hypothesis must explain the facts that gave rise to the need for explanation. This means that by using the hypothesis plus other known and accepted generalizations, one should be able to deduce the original problem condition. Thus hypothesis must actually explain what it claims to explain; it should have empirical reference.

**TESTING OF HYPOTHESIS:**

As has been stated above that hypothesis testing determines the validity of the assumption (technically described as null hypothesis) with a view to choose between two conflicting hypotheses about the value of a population parameter. Hypothesis testing helps to decide on the basis of a sample data, whether a hypothesis about the population is likely to be true or false. Statisticians have developed several tests of hypotheses (also known as the tests of significance) for the purpose of testing of hypotheses which can be classified as:

1. Parametric tests or standard tests of hypotheses; and
2. Non-parametric tests or distribution-free test of hypothesis.

Parametric tests usually assume certain properties of the parent population from which we draw samples. Assumptions like observations come from a normal population, sample size is large, assumptions about the population parameters like mean, variance, etc., must hold good before parametric tests can be used. But there are situations when the researcher cannot or does not want to make such assumptions. In such situations we use statistical methods for testing hypotheses which are called non-parametric tests because such tests do not depend on any assumption about the parameters of the parent population. Besides, most non-parametric tests assume only nominal or ordinal data, whereas parametric tests require measurement equivalent to at least an interval scale.

**UNIT – IV**

**STATISTICAL TECHNIQUES**

**Introduction:** Statistical methods involved in carrying out a study include planning, designing, collecting data, analyzing, drawing meaningful interpretation and reporting of the research findings. The statistical analysis gives meaning to the meaningless numbers, thereby breathing life into a lifeless data. The results and inferences are precise only if proper statistical tests are used. This article will try to acquaint the reader with the basic research tools that are utilized while conducting various studies. The article covers a brief outline of the variables, an understanding of quantitative and qualitative variables and the measures of central tendency. An idea of the sample size estimation, power analysis and the statistical errors is given. Finally, there is a summary of parametric and non-parametric tests used for data analysis.

**Key words:**Basic statistical tools, degree of dispersion, measures of central tendency, parametric tests and non-parametric tests, variables, variance

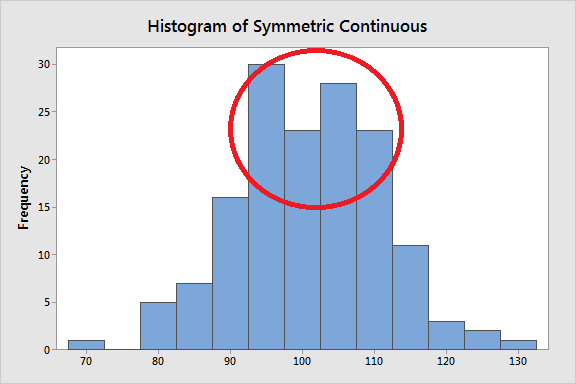
**Meaning:** Statistics is a branch of science that deals with the collection, organisation, analysis of data and drawing of inferences from the samples to the whole population. This requires a proper design of the study, an appropriate selection of the study sample and choice of a suitable statistical test. An adequate knowledge of statistics is necessary for proper designing of an epidemiological study or a clinical trial. Improper statistical methods may result in erroneous conclusions which may lead to unethical practice.

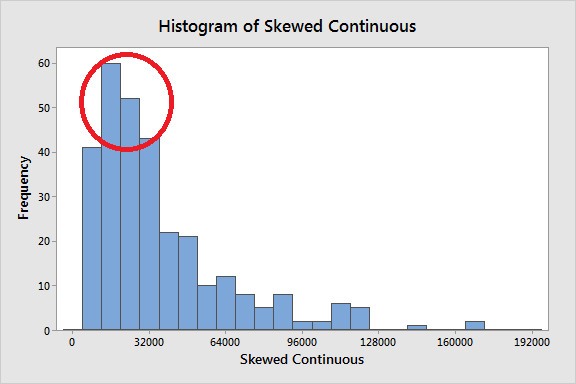
**MEASURE OF CENTRAL TENDANCY**

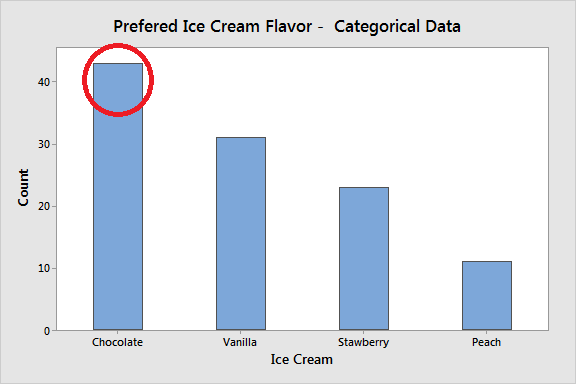
**Meaning:** A measure of central tendency is a summary statistic that represents the center point or typical value of a dataset. These measures indicate where most values in a distribution fall and are also referred to as the central location of a distribution. You can think of it as the tendency of data to cluster around a middle value. In [statistics](https://statisticsbyjim.com/glossary/statistics/), the three most common measures of central tendency are the [mean](https://statisticsbyjim.com/glossary/mean/), [median](https://statisticsbyjim.com/glossary/median/), and [mode](https://statisticsbyjim.com/glossary/mode/). Each of these measures calculates the location of the central point using a different method.

Choosing the best measure of central tendency depends on the type of data you have. In this post, I explore these measures of central tendency; show you how to calculate them, and how to determine which one is best for your data.

## EXAMPLES: Most articles that you’ll read about the mean, median and mode focus on how you calculate each one. I’m going to take a slightly different approach to start out. My philosophy throughout my blog is to help you intuitively grasp statistics by focusing on concepts. Consequently, I’m going to start by illustrating the central point of several datasets graphically—so you understand the goal. Then, we’ll move on to choosing the best measure of central tendency for your data and the calculations.

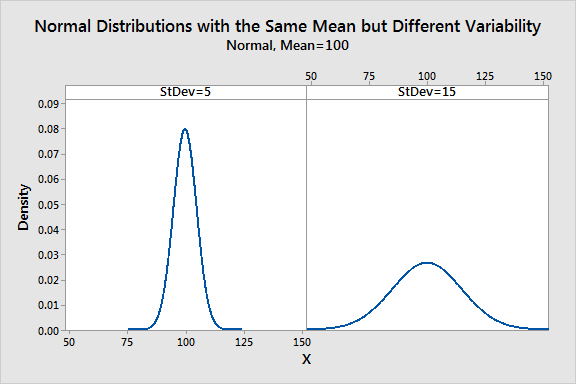
The three distributions below represent different data conditions. In each distribution, look for the region where the most common values fall. Even though the shapes and type of data are different, you can find that central location. That’s the area in the distribution where the most common values are located.





As the graphs highlight, you can see where most values tend to occur. That’s the concept. Measures of central tendency represent this idea with a value. Coming up, you’ll learn that as the distribution and kind of data changes, so does the best measure of central tendency. Consequently, you need to know the type of data you have, and graph it, before choosing a measure of central tendency!.

The central tendency of a distribution represents one characteristic of a distribution. Another aspect is the variability around that central value. While measures of variability is the topic of a different article (link below), this property describes how far away the data points tend to fall from the center. The graph below shows how distributions with the same central tendency (mean = 100) can actually be quite different. The panel on the left displays a distribution that is tightly clustered around the mean, while the distribution on the right is more spread out. It is crucial to understand that the central tendency summarizes only one aspect of a distribution and that it provides an incomplete picture by itself.

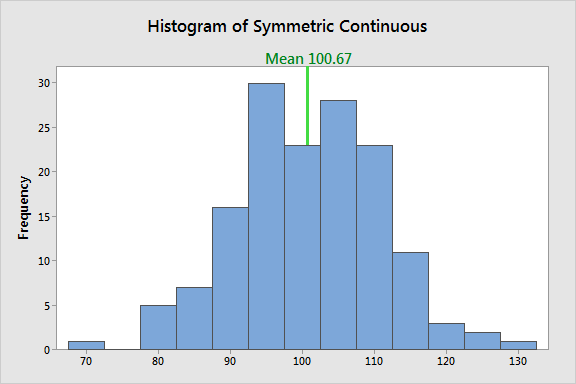


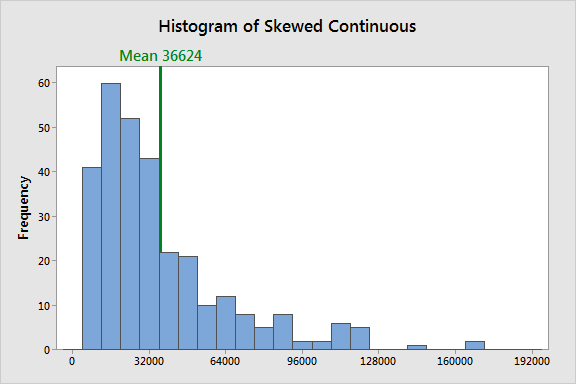
**Related event**: [Measures of Variability: Range, Interquartile Range, Variance, and Standard Deviation](https://statisticsbyjim.com/basics/variability-range-interquartile-variance-standard-deviation/)

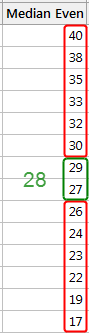
## Mean.

The mean is the arithmetic [average](https://statisticsbyjim.com/glossary/mean/), and it is probably the measure of central tendency that you are most familiar. Calculating the mean is very simple. You just add up all of the values and divide by the number of observations in your dataset.

{\displaystyle \frac {x_{1}+x_{2}+\cdots +x_{n}}{n}}The calculation of the mean incorporates all values in the data. If you change any value, the mean changes. However, the mean doesn’t always locate the center of the data accurately. Observe the histograms below where I display the mean in the distributions.







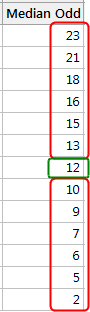
In a symmetric distribution, the mean locates the center accurately.

However, in a [skewed](https://statisticsbyjim.com/glossary/skewed-data/) distribution, the mean can miss the mark. In the histogram above, it is starting to fall outside the central area. This problem occurs because [outliers](https://statisticsbyjim.com/glossary/outliers/) have a substantial impact on the mean. Extreme values in an extended tail pull the mean away from the center. As the distribution becomes more skewed, the mean is drawn further away from the center. Consequently, it’s best to use the mean as a measure of the central tendency when you have a symmetric distribution.

**When to use the mean**: Symmetric distribution, [Continuous data](https://statisticsbyjim.com/glossary/continuous-variables/).

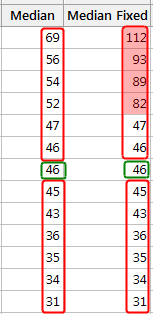
## Median: The median is the middle value. It is the value that splits the dataset in half. To find the median, order your data from smallest to largest, and then find the data point that has an equal amount of values above it and below it. The method for locating the median varies slightly depending on whether your dataset has an even or odd number of values. I’ll show you how to find the median for both cases. In the examples below, I use whole numbers for simplicity, but you can have decimal places.

In the dataset with the odd number of observations, notice how the number 12 has six values above it and six below it. Therefore, 12 is the median of this dataset.



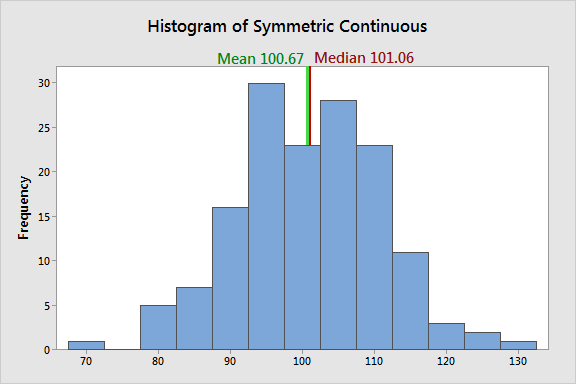
When there is an even number of values, you count in to the two innermost values and then take the average. The average of 27 and 29 is 28. Consequently, 28 is the median of this dataset.

[Outliers](https://statisticsbyjim.com/glossary/outliers/) and [skewed data](https://statisticsbyjim.com/glossary/skewed-data/) have a smaller [effect](https://statisticsbyjim.com/glossary/effect/) on the median. To understand why, imagine we have the [Median](https://statisticsbyjim.com/glossary/median/) dataset below and find that the median is 46. However, we discover data entry errors and need to change four values, which are shaded in the Median Fixed dataset. We’ll make them all significantly higher so that we now have a skewed distribution with large outliers.

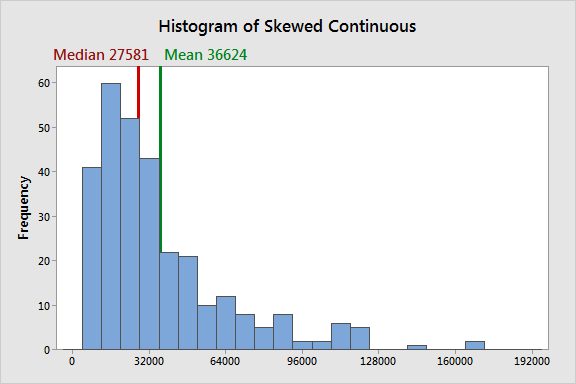


As you can see, the median doesn’t change at all. It is still 46. Unlike the mean, the median value doesn’t depend on all the values in the dataset. Consequently, when some of the values are more extreme, the effect on the median is smaller. Of course, with other types of changes, the median can change. When you have a skewed distribution, the median is a better measure of central tendency than the mean.

### Comparing the mean and median: Now, let’s test the median on the symmetrical and skewed distributions to see how it performs, and I’ll include the mean on the histograms so we can make comparisons.



In a symmetric distribution, the mean and median both find the center accurately. They are approximately equal.



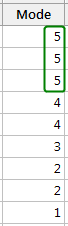
In a skewed distribution, the outliers in the tail pull the mean away from the center towards the longer tail. For this example, the mean and median differ by over 9000, and the median better represents the central tendency for the distribution.

These data are based on the U.S. household income for 2006. Income is the classic example of when to use the median because it tends to be skewed. The median indicates that half of all incomes fall below 27581, and half are above it. For these data, the mean overestimates where most household incomes fall.

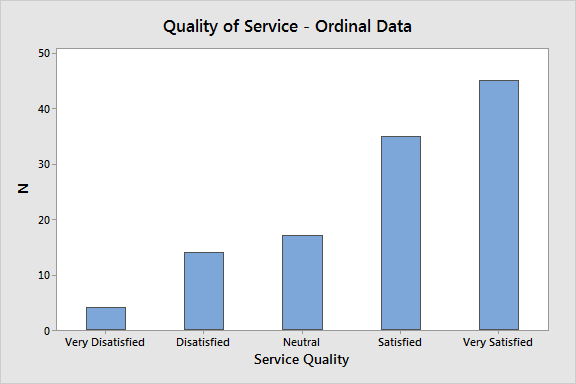
**When to use the median**: [Skewed](https://statisticsbyjim.com/glossary/skewed-data/) distribution, Continuous data, [Ordinal data](https://statisticsbyjim.com/glossary/ordinal-variables/)

## Mode: The mode is the value that occurs the most frequently in your data set. On a bar chart, the mode is the highest bar. If the data have multiple values that are tied for occurring the most frequently, you have a multimodal distribution. If no value repeats, the data do not have a mode.

In the dataset below, the value 5 occurs most frequently, which makes it the mode. These data might represent a 5-point Likert scale.



Typically, you use the mode with categorical, ordinal, and discrete data. In fact, the mode is the only measure of central tendency that you can use with [categorical data](https://statisticsbyjim.com/glossary/categorical-variables/)—such as the most preferred flavor of ice cream. However, with categorical data, there isn’t a central value because you can’t order the groups. With ordinal and discrete data, the mode can be a value that is not in the center. Again, the mode represents the most common value.



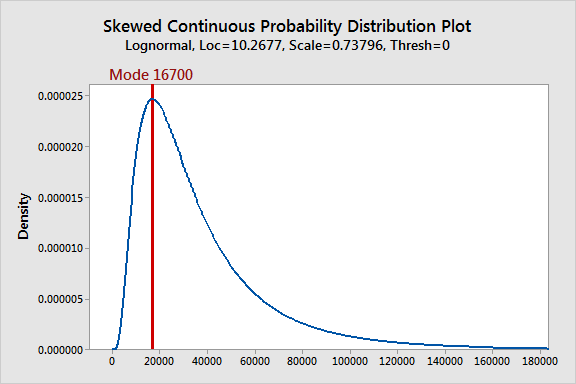
In the graph of service quality, Very Satisfied is the mode of this distribution because it is the most common value in the data. Notice how it is at the extreme end of the distribution. I’m sure the service providers are pleased with these results!

### Continuous data where no values repeat. This dataset does not have a mode.Finding the mode for continuous data

In the [continuous data](https://statisticsbyjim.com/glossary/continuous-variables/) below, no values repeat, which means there is no mode. With continuous data, it is unlikely that two or more values will be exactly equal because there are an infinite number of values between any two values.

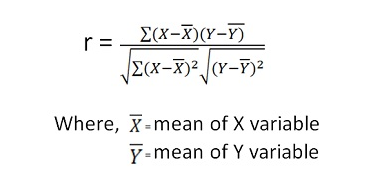
When you are working with the raw continuous data, don’t be surprised if there is no mode. However, you can find the mode for continuous data by locating the maximum value on a probability distribution plot. If you can [identify a probability distribution that fits your data](https://statisticsbyjim.com/hypothesis-testing/identify-distribution-data/), find the peak value and use it as the mode.

The probability distribution plot displays a lognormal distribution that has a mode of 16700. This distribution corresponds to the U.S. household income example in the median section.



Karl Pearson’s coefficient of correlation: A correlation coefficient is generally applied in statistics to calculate a relationship between two variables. The correlation shows a specific value of a degree of a linear relationship between X and Y variables. There are various types of a correlation coefficient, but Pearson’s correlation (also called Pearson’s R) is the correlation coefficient frequently used in linear regression.

Pearson’s Coefficient Correlation: Karl Pearson’s Coefficient of Correlation is an extensively used mathematical method in which the numerical representation is applied to measure the level of relation between linear related variables. The coefficient of correlation is expressed by **“r”.**

**Karl Pearson Correlation Coefficient Formula**

Alternative Formula (covariance formula)

https://cdn1.byjus.com/wp-content/uploads/2019/06/word-image29.png

**Pearson correlation example**

* When a correlation coefficient is (1) that means every increase in one variable, there is a positive increase in other fixed proportion. For instance, shoe sizes change according to the length of the foot and are (almost) perfect correlation.
* When a correlation coefficient is (-1) that means every positive increase in one variable, there is a negative decrease in other fixed proportion. For instance, with the decrease in the quantity of gas in a gas tank, it shows (almost) a perfect correlation with speed.
* When a correlation coefficient is (0) for every increase, it means there is no positive or negative increase and the two variables are not related.

**ACTUAL MEAN METHOD**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Q.1 FROM THE FOLLOWING DATA COMPUTE KARL PERASON’S COEFFICIENT OF CORRELATION. (USING ACTUAL MEAN METHOD).** | | | | | | | | | | | | |
| **Price (`)** | | 10 | | 20 | | 30 | 40 | | 50 | 60 | | 70 |
| **Supply (Units)** | | 8 | | 6 | | 14 | 16 | | 10 | 20 | | 24 |
| **Q.2 FROM THE FOLLOWING DATA COMPUTE KARL PERASON’S COEFFICIENT OF CORRELATION. (USE ACTUAL MEAN METHOD).** | | | | | | | | | | | | |
| **X** | 15 | | 18 | | 20 | | | 28 | | | 34 | |
| **Y** | 40 | | 42 | | 46 | | | 50 | | | 52 | |

**ASSUMED MEAN METHOD**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Q.1 FROM THE FOLLOWING DATA COMPUTE KARL PERASON’S COEFFICIENT OF CORRELATION. (USING ASSUMED MEAN METHOD).** | | | | | | | | | | | | | | | |
| **Price (`)** | 10 | | 20 | | 30 | | | 40 | | 50 | | 60 | | 70 | |
| **Supply (Units)** | 8 | | 6 | | 14 | | | 16 | | 10 | | 20 | | 24 | |
| **Q.2 FROM THE FOLLOWING DATA COMPUTE CORRELATION BETWEEN HEIGHT OF FATHER AND HEIGHT OF DAUGHTERS BY KARL PERASON’S COEFFICIENT OF CORRELATION. (USING ASSUMED MEAN METHOD).** | | | | | | | | | | | | | | | |
| **Height of Father (Cms)** | | 65 | | 66 | | 67 | 67 | | 68 | | 69 | | 71 | | 73 |
| **Height of Daughter (Cms)** | | 67 | | 68 | | 64 | 69 | | 72 | | 70 | | 69 | | 73 |

**STEP DEVIATION METHOD**

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Q.1 FROM THE FOLLOWING DATA COMPUTE KARL PERASON’S COEFFICIENT OF CORRELATION. (USING STEP DEVIATION METHOD).** | | | | | | | | | | | | | |
| **Price (`)** | 10 | | 20 | | 30 | | 40 | | 50 | | 60 | | 70 |
| **Supply (Units)** | 8 | | 6 | | 14 | | 16 | | 10 | | 20 | | 24 |
| **Q.2 FROM THE FOLLOWING DATA COMPUTE KARL PERASON’S COEFFICIENT OF CORRELATION. (USE STEP DEVIATION METHOD).** | | | | | | | | | | | | | |
| **Density (Per Sq. Km)** | | 2000 | | 5000 | | 4000 | | 7000 | | 6000 | | 3000 | |
| **Patients of dengue fever** | | 100 | | 160 | | 140 | | 200 | | 170 | | 130 | |

**DIRECT METHOD**

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Q.1 FROM THE FOLLOWING DATA COMPUTE KARL PERASON’S COEFFICIENT OF CORRELATION. (USING DIRECT METHOD).** | | | | | | | | | | | | |
| **Price (`)** | 10 | 20 | | 30 | | 40 | | 50 | | 60 | | 70 |
| **Supply (Units)** | 8 | 6 | | 14 | | 16 | | 10 | | 20 | | 24 |
| **Q.2 FROM THE FOLLOWING DATA COMPUTE KARL PERASON’S COEFFICIENT OF CORRELATION. (USING DIRECT METHOD).** | | | | | | | | | | | | |
| **Price (in `)** | | | 5 | | 6 | | 3 | | 4 | | 3 | |
| **Demand (in Units)** | | | 10 | | 10 | | 12 | | 11 | | 12 | |

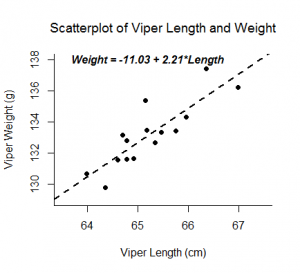
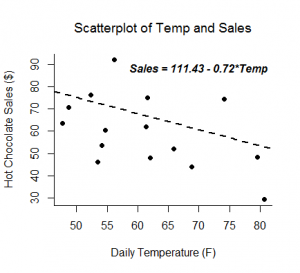
**REGRESSION**

A linear regression analysis produces estimates for the **slope**and **intercept**of the linear equation predicting an outcome variable, Y, based on values of a predictor variable, X.  A general form of this equation is shown below:

[regression equation](http://sites.utexas.edu/sos/files/2015/07/regression-equation.png)

The intercept, b0,  is the predicted value of Y when X=0.  The slope, b1, is the average change in Y for every one unit increase in X.  Beyond giving you the strength and direction of the linear relationship between X and Y, the slope estimate allows an interpretation for how Y changes when X increases. This equation can also be used to predict values of Y for a value of X.

Examples:

[](http://sites.utexas.edu/sos/files/2015/07/Scatter_line_vipers.png)       [](http://sites.utexas.edu/sos/files/2015/07/Scatter_line_temp.png)

Inferential tests can be run on both the correlation and slope estimates calculated from a random sample from a population. Both analyses are t-tests run on the null hypothesis that the two variables are not linearly related. If run on the same data, a correlation test and slope test provide the same test statistic and p-value.

* [Random samples](http://sites.utexas.edu/sos/random/)
* [Independent observations](http://sites.utexas.edu/sos/indobs/)
* The predictor variable and outcome variable are linearly related (assessed by visually checking a scatterplot).
* The population of values for the outcome are normally distributed for each value of the predictor (assessed by confirming the [normality](http://sites.utexas.edu/sos/normal/) of the residuals).
* The variance of the distribution of the outcome is the same for all values of the predictor (assessed by visually checking a residual plot for a funneling pattern).

**Hypotheses:**

Ho: The two variables are not linearly related.  
Ha: The two variables are linearly related.

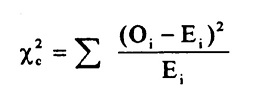
## What is a Chi Square Test?

There are **two types of chi-square tests**. Both use the chi-square statistic and distribution for different purposes:

* A**chi-square goodness of fit test** determines if a sample data matches a population. For more details on this type, see: [Goodness of Fit Test](https://www.statisticshowto.com/goodness-of-fit-test/).
* A **chi-square test for independence** compares two variables in a contingency table to see if they are related. In a more general sense, it tests to see whether distributions of [categorical variables](https://www.statisticshowto.com/what-is-a-categorical-variable/) differ from each another.
  + A **very small chi square test statistic** means that your observed data fits your expected data extremely well. In other words, there is a relationship.
  + A **very large chi square test statistic**means that the data does not fit very well. In other words, there isn’t a relationship.

## What is a Chi-Square Statistic?

The formula for the chi-square statistic used in the chi square test is:

[](https://www.statisticshowto.com/wp-content/uploads/2013/09/chi-square-formula.jpg)

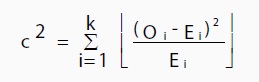
**The chi-square formula.**

The subscript “c” are the degrees of freedom. “O” is your observed value and E is your [expected value](https://www.statisticshowto.com/probability-and-statistics/expected-value/). It’s very rare that you’ll want to actually use this formula to find a critical chi-square value by hand. The summation symbol means that you’ll have to perform a calculation for every single data item in your data set. As you can probably imagine, the calculations can get very, very, lengthy and tedious. Instead, you’ll probably want to use technology:

* [Chi Square Test in SPSS.](https://www.statisticshowto.com/probability-and-statistics/chi-square/#SPSSChi)
* [Chi Square P-Value in Excel.](https://www.statisticshowto.com/calculate-chi-square-p-value-excel/)

A chi-square statistic is one way to show a relationship between two [categorical variables](https://www.statisticshowto.com/what-is-a-categorical-variable/). In statistics, there are two types of variables: numerical (countable) variables and non-numerical (categorical) variables. The chi-squared statistic is a single number that tells you how much difference exists between your observed counts and the counts you would expect if there were no relationship at all in the population.

There are a **few variations** on the chi-square statistic. Which one you use depends upon how you collected the data and which hypothesis is being tested. However, all of the variations use the same idea, which is that you are comparing your expected values with the values you actually collect. One of the most common forms can be used for [contingency tables](https://www.statisticshowto.com/what-is-a-contingency-table/):

[](https://www.statisticshowto.com/wp-content/uploads/2013/07/chi-square.jpg)

Where O is the observed value, E is the expected value and “i” is the “ith” position in the contingency table.

A **low value** for chi-square means there is a high correlation between your two sets of data. In theory, if your observed and expected values were equal (“no difference”) then chi-square would be zero — an event that is unlikely to happen in real life. Deciding whether a chi-square test statistic is large enough to indicate a [statistically significant](https://www.statisticshowto.com/what-is-statistical-significance/) difference isn’t as easy it seems. It would be nice if we could say a chi-square test statistic >10 means a difference, but unfortunately that isn’t the case.

You could take your calculated chi-square value and compare it to a [critical value from a chi-square table.](https://www.statisticshowto.com/tables/chi-squared-table-right-tail/) If the chi-square value is more than the critical value, then there is a significant difference.

You could also use a [p-value.](https://www.statisticshowto.com/p-value/) First state the [null hypothesis](https://www.statisticshowto.com/probability-and-statistics/null-hypothesis/#state) and the [alternate hypothesis](https://www.statisticshowto.com/what-is-an-alternate-hypothesis/). Then generate a chi-square curve for your results along with a p-value (See: [Calculate a chi-square p-value Excel](https://www.statisticshowto.com/calculate-chi-square-p-value-excel/)). Small p-values (under 5%) usually indicate that a difference is significant (or “small enough”).

**Tip**: The Chi-square statistic can only be used on numbers. They can’t be used for percentages, proportions, means or similar statistical value. For example, if you have 10 percent of 200 people, you would need to convert that to a number (20) before you can run a test statistic.

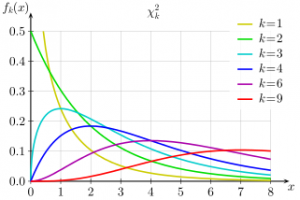
## Chi Square P-Values: A chi square test will give you a [p-value](https://www.statisticshowto.com/p-value/). The p-value will tell you if your test results are [significant](https://www.statisticshowto.com/what-is-statistical-significance/)or not. In order to perform a chi square test and get the p-value, you need two pieces of information:

1. [Degrees of freedom](https://www.statisticshowto.com/degrees-of-freedom/). That’s just the number of categories minus 1.
2. The [alpha level](https://www.statisticshowto.com/what-is-an-alpha-level/) (α). This is chosen by you, or the researcher. The usual alpha level is 0.05 (5%), but you could also have other levels like 0.01 or 0.10.

In elementary statistics or [AP statistics](http://apcentral.collegeboard.com/apc/public/courses/teachers_corner/2151.html), both the degrees of freedom (df) and the alpha level are usually given to you in a question. You don’t normally have to figure out what they are. You may have to figure out the df yourself, but it’s pretty simple: count the categories and subtract 1.

**Degrees of freedom** are placed as a [subscript](http://www.merriam-webster.com/dictionary/subscript)after the chi-square (Χ2) symbol. For example, the following chi square shows 6 df:  
Χ26.  
And this chi square shows 4 df:  
Χ24.

## The Chi-Square Distribution

[](https://www.statisticshowto.com/wp-content/uploads/2014/11/chi-square-distribution.png)

*By Geek3|Wikimedia Commons*[*GFDL*](https://www.gnu.org/copyleft/fdl.html)

The chi-square distribution (also called the chi-squared distribution) is a special case of the [gamma distribution](https://www.statisticshowto.com/gamma-distribution/); A chi square distribution with n [degrees of freedom](https://www.statisticshowto.com/degrees-of-freedom/) is equal to a gamma distribution with a = n / 2 and b = 0.5 (or β = 2).

Let’s say you have a [random sample](https://www.statisticshowto.com/simple-random-sample/) taken from a [normal distribution](https://www.statisticshowto.com/probability-and-statistics/normal-distributions/). The chi square distribution is the distribution of the sum of these random samples **squared**. The **degrees of freedom (k)**are equal to the number of [samples](https://www.statisticshowto.com/sample/)being summed. For example, if you have taken 10 samples from the normal distribution, then df = 10. The degrees of freedom in a chi square distribution is also its [**mean**](https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/mean-median-mode/#mean). In this example, the mean of this particular distribution will be 10. Chi square distributions are always [right skewed](https://www.statisticshowto.com/probability-and-statistics/skewed-distribution/#SkewRight). However, the greater the degrees of freedom, the more the chi square distribution looks like a normal distribution.

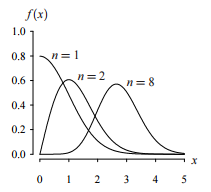
**The chi-squared distribution has many uses in statistics, including:**

* [Confidence interval](https://www.statisticshowto.com/probability-and-statistics/confidence-interval/)estimation for a population [standard deviation](https://www.statisticshowto.com/probability-and-statistics/standard-deviation/) of a normal distribution from a sample standard deviation.
* Independence of two criteria of classification of [qualitative variables](https://www.statisticshowto.com/qualitative-variable/).
* Relationships between [categorical variables](https://www.statisticshowto.com/what-is-a-categorical-variable/) ([contingency tables](https://www.statisticshowto.com/what-is-a-contingency-table/)).
* [Sample variance](https://www.statisticshowto.com/probability-and-statistics/descriptive-statistics/sample-variance/) study when the underlying distribution is normal.
* Tests of deviations of differences between expected and observed frequencies (one-way tables).
* The chi-square test (a [goodness of fit](https://www.statisticshowto.com/goodness-of-fit-test/)test).

## Chi Distribution

A similar distribution is the **chi distribution**. This distribution describes the **square root**of a variable distributed according to a chi-square distribution. With df = n > 0 [degrees of freedom](https://www.statisticshowto.com/degrees-of-freedom/) has a [probability density function](https://www.statisticshowto.com/probability-density-function/) of:

*f(x) = 2(1-n/2) x(n-1) e(-(x2)/2) / Γ(n/2)*

For values where x is positive.  
[](https://www.statisticshowto.com/wp-content/uploads/2013/07/chi-distribution.png)

**ANOVA**

An **ANOVA**test is a way to find out if survey or experiment results are [significant](https://www.statisticshowto.com/what-is-statistical-significance/). In other words, they help you to figure out if you need to [reject the null hypothesis](https://www.statisticshowto.com/support-or-reject-null-hypothesis/) or accept the [alternate hypothesis](https://www.statisticshowto.com/what-is-an-alternate-hypothesis/). Basically, **you’re testing groups to see if there’s a difference between them.** Examples of when you might want to test different groups:

* A group of psychiatric patients are trying three different therapies: counseling, medication and biofeedback. You want to see if one therapy is better than the others.
* A manufacturer has two different processes to make light bulbs. They want to know if one process is better than the other.
* Students from different colleges take the same exam. You want to see if one college outperforms the other.

## What Does “One-Way” or “Two-Way Mean?

**One-way** or **two-way** refers to the number of [independent variables](https://www.statisticshowto.com/independent-variable-definition/) (IVs) in your Analysis of Variance test.

* One-way has one independent variable (with 2 [levels](https://www.statisticshowto.com/levels-in-statistics/)). For example: brand of cereal,
* Two-way has two independent variables (it can have multiple levels). For example: brand of cereal, calories.

## What are “Groups” or “Levels”?

Groups or levels are different groups within the same [independent variable](https://www.statisticshowto.com/independent-variable-definition/). In the above example, your levels for “brand of cereal” might be Lucky Charms, Raisin Bran, Cornflakes — a total of three levels. Your levels for “Calories” might be: sweetened, unsweetened — a total of two levels.

Let’s say you are studying if an alcoholic support group and individual counseling combined is the most effective treatment for lowering alcohol consumption. You might split the study participants into three groups or levels:

* Medication only,
* Medication and counseling,
* Counseling only.

Your [dependent variable](https://www.statisticshowto.com/dependent-variable-definition/) would be the number of alcoholic beverages consumed per day.

If your groups or levels have a hierarchical structure (each level has unique subgroups), then use a [nested ANOVA](https://www.statisticshowto.com/nested-model-anova-factors/#anova) for the analysis.

## What Does “Replication” Mean?

It’s whether you are replicating (i.e. duplicating) your test(s) with multiple groups. With a two way ANOVA with replication, you have two groups and individuals within that group are doing more than one thing (i.e. two groups of students from two colleges taking two tests). If you only have one group taking two tests, you would use **without replication.**

## Types of Tests.

There are two main types: one-way and two-way. Two-way tests can be with or without replication.

* One-way ANOVA between groups: used when you want to test **two groups** to see if there’s a difference between them.
* Two way ANOVA without replication: used when you have **one group** and you’re**double-testing**that same group. For example, you’re testing one set of individuals before and after they take a medication to see if it works or not.
* Two way ANOVA with replication: **Two groups**, and the members of those groups are **doing more than one thing**. For example, two groups of patients from different hospitals trying two different therapies.

## One Way ANOVA

A one way ANOVA is used to compare two means from two independent (unrelated) groups using the [F-distribution](https://www.statisticshowto.com/probability-and-statistics/f-statistic-value-test/#Fdist). The [null hypothesis](https://www.statisticshowto.com/probability-and-statistics/null-hypothesis/) for the test is that the two [means](https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/mean-median-mode/#mean)are equal. Therefore, a [significant](https://www.statisticshowto.com/what-is-statistical-significance/)result means that the two means are unequal.

## Examples of when to use a one way ANOVA

**Situation 1:** You have a group of individuals randomly split into smaller groups and completing different tasks. For example, you might be studying the effects of tea on weight loss and form three groups: green tea, black tea, and no tea.

**Situation 2:** Similar to situation 1, but in this case the individuals are split into groups based on an attribute they possess. For example, you might be studying leg strength of people according to weight. You could split participants into weight categories (obese, overweight and normal) and measure their leg strength on a weight machine.

## Limitations of the One Way ANOVA

A one way ANOVA will tell you that at least two groups were different from each other. But **it won’t tell you which groups were different.** If your test returns a significant f-statistic, you may need to run an [ad hoc test](https://amzn.to/2sQXcxi) (like the [Least Significant Difference](https://www.statisticshowto.com/how-to-calculate-the-least-significant-difference-lsd/) test) to tell you exactly which groups had a [difference in means](https://www.statisticshowto.com/mean-difference/).

## Two Way ANOVA

A Two Way ANOVA is an extension of the One Way ANOVA. With a One Way, you have one [independent variable](https://www.statisticshowto.com/independent-variable-definition/) affecting a [dependent variable](https://www.statisticshowto.com/dependent-variable-definition/). With a Two Way ANOVA, there are two independents. Use a two way ANOVA when you have one [measurement variable](https://www.statisticshowto.com/quantitative-variables-data/) (i.e. a [quantitative variable](https://www.statisticshowto.com/what-are-quantitative-variables-and-quantitative-data/)) and two [nominal variables](https://www.statisticshowto.com/nominal-ordinal-interval-ratio/). In other words, if your experiment has a quantitative outcome and you have two categorical [explanatory variables](https://www.statisticshowto.com/explanatory-variable/), a two way ANOVA is appropriate.

For example, you might want to find out if there is an interaction between income and gender for anxiety level at job interviews. The anxiety level is the outcome, or the variable that can be measured. Gender and Income are the two [categorical variables](https://www.statisticshowto.com/what-is-a-categorical-variable/). These categorical variables are also the independent variables, which are called **factors** in a Two Way ANOVA.

The factors can be split into **levels**. In the above example, income level could be split into three levels: low, middle and high income. Gender could be split into three levels: male, female, and transgender. Treatment groups are all possible combinations of the factors. In this example there would be 3 x 3 = 9 treatment groups.

## Main Effect and Interaction Effect

The results from a Two Way ANOVA will calculate a[main effect](https://www.statisticshowto.com/main-effect/) and an [interaction effect](https://www.statisticshowto.com/interaction-effect-interacting-variable/). The main effect is similar to a One Way ANOVA: each factor’s effect is considered separately. With the interaction effect, all factors are considered at the same time. Interaction effects between factors are easier to test if there is more than one observation in each cell. For the above example, multiple stress scores could be entered into cells. If you do enter multiple observations into cells, the number in each cell must be equal.

Two [null hypotheses](https://www.statisticshowto.com/probability-and-statistics/null-hypothesis/) are tested if you are placing one observation in each cell. For this example, those hypotheses would be:

H01: All the income groups have equal mean stress.

H02: All the gender groups have equal mean stress.

**For multiple observations in cells, you would also be testing a third hypothesis:**

H03: The factors are independent or the interaction effect does not exist.

An [F-statistic](https://www.statisticshowto.com/probability-and-statistics/f-statistic-value-test/) is computed for each hypothesis you are testing.

## Assumptions for Two Way ANOVA

* The [population](https://www.statisticshowto.com/what-is-a-population/)must be close to a [normal distribution](https://www.statisticshowto.com/probability-and-statistics/normal-distributions/).
* [Samples](https://www.statisticshowto.com/sample/)must be independent.
* Population [variances](https://www.statisticshowto.com/probability-and-statistics/variance/) must be equal.
* Groups must have equal [sample sizes](https://www.statisticshowto.com/probability-and-statistics/find-sample-size/).

# Spearman's Rank-Order Correlation

This guide will tell you when you should use Spearman's rank-order correlation to analyse your data, what assumptions you have to satisfy, how to calculate it, and how to report it. If you want to know how to run a Spearman correlation in SPSS Statistics, go to our [Spearman's correlation in SPSS Statistics](https://statistics.laerd.com/spss-tutorials/spearmans-rank-order-correlation-using-spss-statistics.php) guide.

## When should you use the Spearman's rank-order correlation?

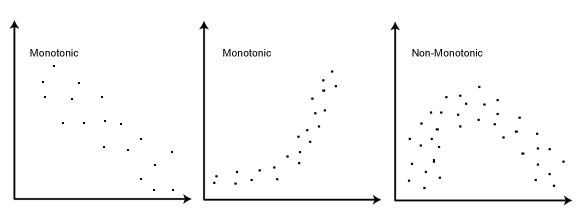
The Spearman's rank-order correlation is the nonparametric version of the [Pearson product-moment correlation](https://statistics.laerd.com/statistical-guides/pearson-correlation-coefficient-statistical-guide.php). Spearman's correlation coefficient, (ρ, also signified by rs) measures the strength and direction of association between two ranked variables.

## What are the assumptions of the test?

You need two variables that are either ordinal, interval or ratio (see our [Types of Variable](https://statistics.laerd.com/statistical-guides/types-of-variable.php) guide if you need clarification). Although you would normally hope to use a Pearson product-moment correlation on interval or ratio data, the Spearman correlation can be used when the assumptions of the Pearson correlation are markedly violated. However, Spearman's correlation determines the strength and direction of the **monotonic relationship** between your two variables rather than the strength and direction of the linear relationship between your two variables, which is what Pearson's correlation determines.

## What is a monotonic relationship?

A monotonic relationship is a relationship that does one of the following: (1) as the value of one variable increases, so does the value of the other variable; or (2) as the value of one variable increases, the other variable value decreases. Examples of monotonic and non-monotonic relationships are presented in the diagram below:



Join the 10,000s of students, academics and professionals who rely on Laerd Statistics. TAKE[PLANS & PRICING](https://statistics.laerd.com/pricing.php)

## Why is a monotonic relationship important to Spearman's correlation?

Spearman's correlation measures the strength and direction of monotonic association between two variables. Monotonicity is "less restrictive" than that of a linear relationship. For example, the middle image above shows a relationship that is monotonic, but not linear.

A monotonic relationship is not strictly an assumption of Spearman's correlation. That is, you can run a Spearman's correlation on a non-monotonic relationship to determine if there is a **monotonic component** to the association. However, you would normally pick a measure of association, such as Spearman's correlation, that fits the pattern of the observed data. That is, if a scatterplot shows that the relationship between your two variables looks monotonic you would run a Spearman's correlation because this will then measure the strength and direction of this monotonic relationship. On the other hand if, for example, the relationship appears linear (assessed via scatterplot) you would run a Pearson's correlation because this will measure the strength and direction of any linear relationship. You will not always be able to visually check whether you have a monotonic relationship, so in this case, you might run a Spearman's correlation anyway.

## How to rank data?

In some cases your data might already be ranked, but often you will find that you need to rank the data yourself (or use [SPSS Statistics](https://statistics.laerd.com/spss-tutorials/ranking-data-in-spss-statistics.php) to do it for you). Thankfully, ranking data is not a difficult task and is easily achieved by working through your data in a table. Let us consider the following example data regarding the marks achieved in a maths and English exam:

|  | Marks | | | | | | | | | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| English | 56 | 75 | 45 | 71 | 61 | 64 | 58 | 80 | 76 | 61 |
| Maths | 66 | 70 | 40 | 60 | 65 | 56 | 59 | 77 | 67 | 63 |

The procedure for ranking these scores is as follows:

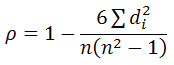
**First, create a table with four columns and label them as below:**

| English (mark) | Maths (mark) | Rank (English) | Rank (maths) |
| --- | --- | --- | --- |
| 56 | 66 | 9 | 4 |
| 75 | 70 | 3 | 2 |
| 45 | 40 | 10 | 10 |
| 71 | 60 | 4 | 7 |
| **61** | 65 | 6.5 | 5 |
| 64 | 56 | 5 | 9 |
| 58 | 59 | 8 | 8 |
| 80 | 77 | 1 | 1 |
| 76 | 67 | 2 | 3 |
| **61** | 63 | 6.5 | 6 |

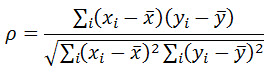
You need to rank the scores for maths and English separately. The score with the highest value should be labelled "1" and the lowest score should be labelled "10" (if your data set has more than 10 cases then the lowest score will be how many cases you have). Look carefully at the two individuals that scored 61 in the English exam (highlighted in bold). Notice their joint rank of 6.5. This is because when you have two identical values in the data (called a "tie"), you need to take the average of the ranks that they would have otherwise occupied. We do this because, in this example, we have no way of knowing which score should be put in rank 6 and which score should be ranked 7. Therefore, you will notice that the ranks of 6 and 7 do not exist for English. These two ranks have been averaged ((6 + 7)/2 = 6.5) and assigned to each of these "tied" scores.

## What is the definition of Spearman's rank-order correlation?

There are two methods to calculate Spearman's correlation depending on whether: (1) your data does not have tied ranks or (2) your data has tied ranks. The formula for when there are no tied ranks is:



where di = difference in paired ranks and n = number of cases. The formula to use when there are tied ranks is:



where i = paired score.

**UNIT – V**

**INTERPRETATION TECHNIQUES**

Trans Cultures provides trainings for interpreters who intend to take professional interpretation certification exams. Below are some of the interpretation techniques we cover in our seminars, and that our interpreters use to interpret for various settings.

#### Consecutive interpretation:

Consecutive interpretation is a technique in which the interpreter serves as an intermediary between the speakers. The interpreter listens to the speech and the speaker pauses and allows the interpreter to repeat their statements in another language. The interpreter begins interpreting after the speaker has paused. Consecutive Interpretation is commonly used for depositions, court hearings, business meetings and negotiations, medical appointments, tours, informal meetings and other social events.

#### Simultaneous interpretation

Simultaneous interpretation is a technique in which the interpreter renders the interpretation simultaneously while the speaker is still speaking. Simultaneous interpretation is used in large conferences, conventions, and seminars where there is only one speaker addressing an audience. The speaker speaks into a microphone and the interpreter wears a headset which enables them to listen to the speaker and almost simultaneously render the message into a microphone in the target language. Those people needing interpretation wear an earpiece which allows them to hear the interpreter. This technique is often used in media interpreting and larger conferences. A simultaneous interpreter must have an absolute command of the target and the source languages, the ability to interpret instantaneously and accurately, and the knowledge of specialized terminology. TransCultures’ interpreters possess all of these skills thanks to our rigorous screening and qualifying process. Many of our conference interpreters are members of prestigious interpreter organizations such as ATA, AIIC and TAALS.

#### Whispered interpretation

Whispered interpretation or Chuchotage is a technique where the interpreter provides interpretation simultaneously to a small audience, usually less than four people. In this setting, the interpreter sits or stands close to the audience and whispers the interpretation into their ears. It is used mainly in bilateral meetings and small settings. Headsets are not needed since it is a small group and no extra time is needed to interpret because the interpretation is done while the speaker is speaking.

#### Sight Translation

Sight translation is a technique that requires the interpreter to render the content of a written document in the source language orally into the target language. This technique of interpretation is often used during legal and medical interpretation.

**RESEARCH REPORT:**

The final step in research is to report the study findings to those who authorized the investigation. The American Marketing association is of the opinion that the findings is marketing research remains valueless unit they are communicated accurately and effectively to the persons who are responsible for policy decisions.

**Meaning of Research Report:**

Research report is the process of communicating the results of an investigation. It is a document which reflects about the research conducted and the care that has been exercised throughout the study.

It is observed that executives are never interested in the methodology adopted by the researcher in the investigation. They are interested only in the final results. From their viewpoint, the research report is the written/oral presentation which the executive see/hear of a project carried out by the researcher but initiated by the decision maker. The research report conveys the information desired by the decision maker in an understandable form.

**According to Lancaster,** “A report is a statement of collected and considered facts, so drawn-up as to give clear and concise information to persons who are not already in possession of the full facts of the subject matter of the report”.

**IMPORTANCE OF RESEARCH REPORT:**

* **Provide Details:** It is the research report gives every other details of research.
* **Source of concise and organized data:** A research report is a published document and as such lucid explanation is to be given for the understanding of every other reader.
* **Logical Presentation:** The purpose of research cannot be served unless it is presented properly.
* **Reflects Final Research:** Skill and care shall be taken to write a report because it is the final work of the research.
* **Tool of Evaluating Researcher:** It is the research report which discloses the scholarliness of the researcher.
* **Bibliographical Evidence:** A research report gives scope of further research and as such it is considered as bibliographical evidence.

**TYPES OF RESEARCH REPORT:**

The results of research instigation can be presented in a number of ways viz.; a technical report, a popular report, an article, a monograph or at times even in the form of oral presentation. Which methods of presentation to be used in a particular study, depends on the circumstances under which the study arose, and the nature of the results.

A technical report is sued whenever a full written report of the study is required whether for record-keeping or for public dissemination. A popular report is used if the research results have policy implications. Below we give a few details about the major types of reports:

* **Technical Report:** In the technical report the main emphasis is on (i) the methods employed, (ii) assumptions made in the course of the study, (iii) the detailed presentation of the findings including their limitations and supporting data.

**A general outline of a technical report can be follows:**

**i) Summary of Results:** A brief view of the main findings just in two or three pages.

**ii) Nature of the Study:** Description of the general objectives of study, formulation of the problem in operational terms, the working hypothesis, the type of analysis and data required etc.

**iii) Research Methodology:** Specific methods used in the study and their limitations. For instance, in sampling studies we should give details of sample design viz., sample size, sample selection, etc.

**iv) Details of Data:** Discussion of data collected; their sources, characteristics and limitations. If secondary data are sued, their suitability to the problem at hand is fully assessed. In case of a survey, the manner in which data were collected should be fully described.

**v) Analysis of Data and Presentation of Findings:** The analysis of data and presentation of findings of the study with supporting data in the form of tables and charts be fully narrated. This, in fact, happens to be the main body of the report usually extending over several chapters.

**vi) Conclusions**: A detailed summary of the findings and the policy implications drawn from the results be explained.

vii) **Bibliography:** Bibliography of various sources consulted be prepared and attached.

**viii) Technical Appendices**: Appendices be given for all technical matters relating to questionnaire, mathematical derivations, elaboration on particular technique of analysis and the like ones.

**ix) Index:** Index must be prepared and be given invariably in the report at the end.

* **Popular Report:** The popular report is one which gives emphasis on simplicity and attractiveness. The simplification should be sought through clear writing, minimization of technical, particularly mathematical, details and liberal use of charts and diagrams.

**We give below a general outline of a popular report:**

**i) Findings and their implications:** Emphasis in the report is given on the findings of most practical interest and on the implications of these findings.

**ii) Recommendations for Follow-up:** Recommendations for action on the basis of the findings of the study is made in this section of the report.

**iii) Objective of the Study:** A general review of how the problem arises is presented along with the specific objectives of the project under study.

**iv) Methods Employed:** A brief and non-technical description of the methods and techniques used, including a short review of the data on which the study is based, is given in this part of the report.

**v) Results:** This section constitutes the main body of the report where in the results of the study are presented in clear and non-technical terms with liberal use of all sorts of illustrations such as charts, diagrams and the like ones.

**vi) Technical Appendices**: More detailed information on methods used, forms, etc. is presented in the forms of appendices. But the appendices are often not detailed if the report is entirely meant for general public.

* **Interim Report:** Where there is a long time lag between data collection and the presentation of the results in the case of a sponsored project, the study may lose its significance and usefulness and sponsor may also lose interest in it. One of the most effective ways to avoid such eventualities is to present an interim report.
* **Summary Report:** A summary report is generally prepared for the consumption of the lay audience Viz., the general public. The preparation of this type of report is desirable for any study whose findings are of general interest. It is written in non-technical, simple language with a liberal use of pictorial charts.
* **Algorithmic Research Report:** There are problems, viz., production scheduling, JIT, supply chain management, line balancing, layout design, portfolio management, etc., exist in reality. The solution for each of the above problems can be obtained through algorithms. So, the researchers should come out with newer algorithms or improved algorithms for such problems.

**PRINCIPLES OF REPORT WRITING:**

* **Easy to Follow:** The body of the report should be written in a self-evident and easy from. Every subject should be presented under different heads. There may be different sub-topics and sub-headings under one heading. The statement may be made in short and relevant paragraphs.

* **Adhere to the study Objectives:** A good report seeks to achieve certain results consisting primarily of answering the questions that derive from the statement of objectives. Merely to report the findings without reference to the objectives is to produce a sterile piece of writing. This requires researchers to be on intimate terms with the problem.
* **Be Selective:** The research report should be comprehensive as well as concise. The report should be long enough to cover all the components and objectives of the research project. But this does not mean that it should be too detailed and wide. It should not be too detailed unless it is really required to be extensive.
* **Be Objective:** Writers must at all times retain their objectivity. Often researchers will be come so enamored of a study that they overlook their scientific role. This is a natural temptation since much marketing research is done within a sales environment.
* **Have a Purposeful Organization:** Mere recording of facts without purpose or organization inevitably results in confusion which leads to loss of interest. The objective of the report is to give the reader the overall “picture” in the shortest possible time. Therefore each paragraph should be written with the thought of its position in the entire report in mind.
* **Write Clearly:** It is easy to say “write clearly.” But this is difficult for most people to accomplish. Clarity has many facets is highly subjective. Despite this subjectivity, some basic principles can be itemized that.

**PREPARATION OF RESEARCH REPORT:**

Having decided on the type of report, the report writer should now concern himself with its preparation. This can best be done when he is clear about what aspects or points are to be covered by it. Let us first consider the format of the research report.

1. **Research Report Format:** A research report can be written in a number of ways. However, three formats are generally followed:

* **Logical Pattern:** The report may use a logical pattern, which implies that the findings are presented in inductive order, i.e., moving from specific to general.
* **Psychological Pattern**: The report may follow a psychological pattern which is almost inverse of the preceding pattern. In this format, the most critical information, i.e., the conclusion, is provided first, after which follow the findings supporting the conclusion.

* **Chronological Format:** The report may use a chronological format wherein information is given along the time dimension, i.e., things which happened earlier precede those which happened later. This form is generally combined with other formats. The chronological format is the least popular though in respect of problems of a historical nature, it is the most appropriate.

1. **Layout Research Report:** There is no one best format for all report. However, the physical format can be employed to create desirable emphasis and clarity. The use of widely spaced, varied margins, separated headings, different type sizes, and colors – all make it possible to emphasize major points and to clarify the sequence and relationship of ideas.

* **Title Page:** The Title page should indicate the topic on which the report has been prepared – the person or agency who has prepared it, the person or agency for which it has been prepared and the date of submission of the report.
* **Table of Contents:** The table of contents is an outline of the order of the chapter, sections, and sub-sections with their respective pages. If report includes a number of charts, figures, tables, maps, diagrams and graphs etc., a separate table for each category would immediately follow the table of contents.
* **Foreword:** This section serves to introduce the reader to the research project. It should give the background of the problem; the importance of the problem; the various dimensions of the problem; and whether any previous research was done that is pertinent to the specific project being reported.
* **Statement of Objectives:** The specific objectives of the report need to be set forty clearly. The reader must know exactly what the report covers. If the particular project is part of the large problem, it is desirable to state the overall problem and the problem solution process.

* **Methodology:** The purpose of the methodology section is to describe the research procedure. This includes the overall research design, the sampling procedures, the data-collection method, the field methods, and analysis procedure.

i) Research Design

ii) Data-collection Methods:

iii) Sampling

iv) Field work

v) Analysis and Interpretation.

* **Limitations:** The report should also point out the main limitations of the research reported therein. This will be helpful to the reader who can form his own opinion as to how far the results are reliable. In addition, it will be useful to researchers who subsequently undertake a study on the same or a related theme.
* **Findings:** Findings are the results of the study. This section makes up the bulk of the report. It is not just an assortment of statistical tables and charts but an organized narrative f the results. Summary tables and graphic methods of presentation should be used liberally.
* **Conclusions and Recommendations:** Conclusions should be drawn with direct reference to the objectives of the study. The readers should be able to read the objectives. Turn to the conclusions section and find specific conclusions relative to each objective.
* **Appendix:** This is the last section of the report. An appendix gives supplementary information which supports the body of the report but which cannot be given with it. If it is given with the main report, it may distort the focus on the main theme and confuse the reader.
* **Bibliography and Reference:** The bibliography contains the detailed information on books, references, journals and other materials, showing the title of the books / reports used in the preparation of the reports, names of authors, publishers, the year of publication and, if possible, the page numbers.
* **Summary Report:** The report format suggested above does not contain any summary section. This exclusion is deliberate. The summary should not prepared until the full report is written. Once the report is completed, a summary can be prepared quickly and efficiently.

**ALL THE BEST!!!**

**THE END.**