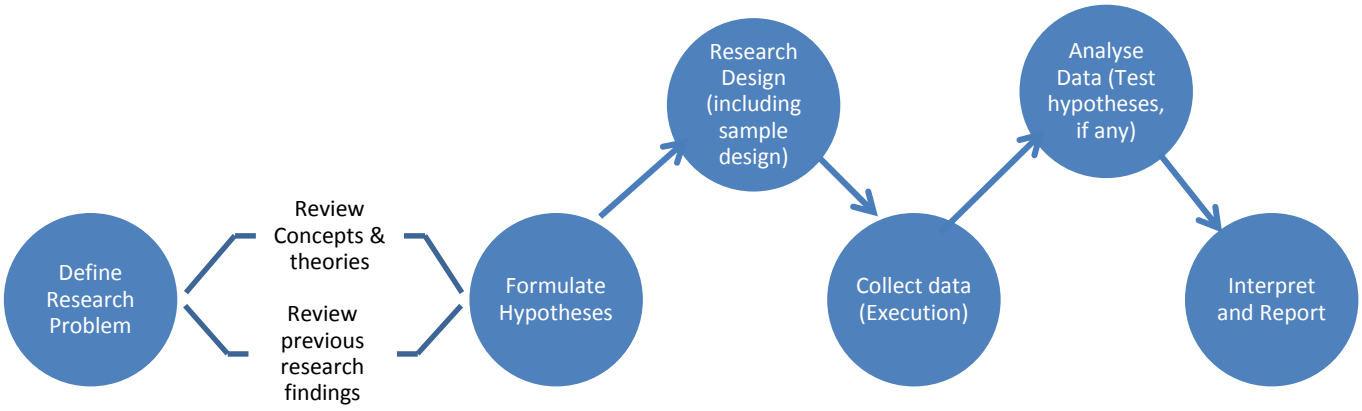
Research Methodology

Definition

Research is exploration for knowledge. It is also defined as a scientific investigation of existing knowledge and/or avenues to innovate solutions for existing challenges.

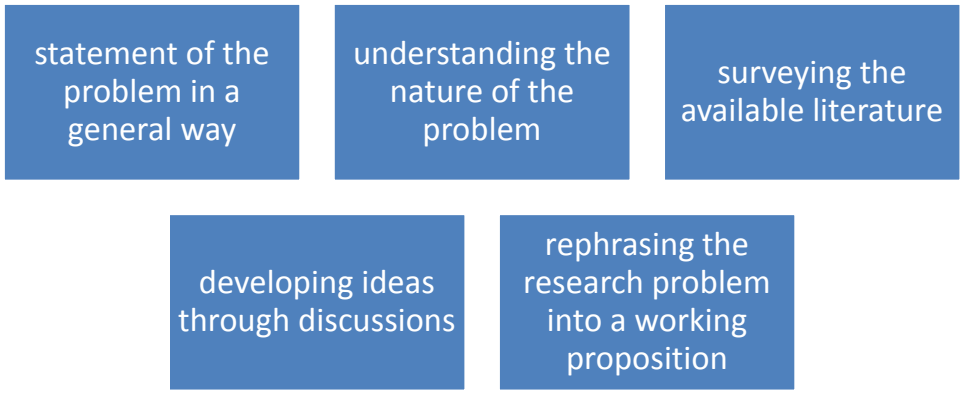
Since, research follows a rationale and a process it is scientific in nature. Methodology of researching into any field is very important as it makes the process more organized and proves as a blueprint of carrying undergoing a research; just like any architect needs a blueprint of the building, similarly, a researcher needs a plan of action, thereby a research methodology.

Research Process



Research Problem:

The topic of research must be precisely stated so that it is easy to understand and also that the researcher has the clarity as to what all should he/she incorporate in his or her research. A research problem should be defined in a systematic manner and should possess least possible ambiguity.



A problem cannot be chosen for the sake of it. The best way to state a research problem is to question yourself about something you believe in. It could be testing a hypotheses whether a given or to be enforced law would be beneficiary to the society? It could be an answer to why India's growth is not as sustainable as China's in terms of secondary sector?

Caution: If the topic's horizon is too broad, it is very likely that the researcher might get confused himself. For instance "India's Growth in 2010"; a researcher feels ambitious to include every possible element of growth and might end up feeling directionless. It is ideal to rephrase it as maybe "India's growth in its Manufacturing sector in 2010".

Criteria for choosing a research problem:

1. Don't choose a topic which is overdone, there is very narrow scope to innovate.
2. The researcher should take into account the feasibility and familiarity of the problem. Also take into consideration the cost, time factor and the one's own knowledge about the topic chosen.

Within explanation of the research of problem, the researcher must state

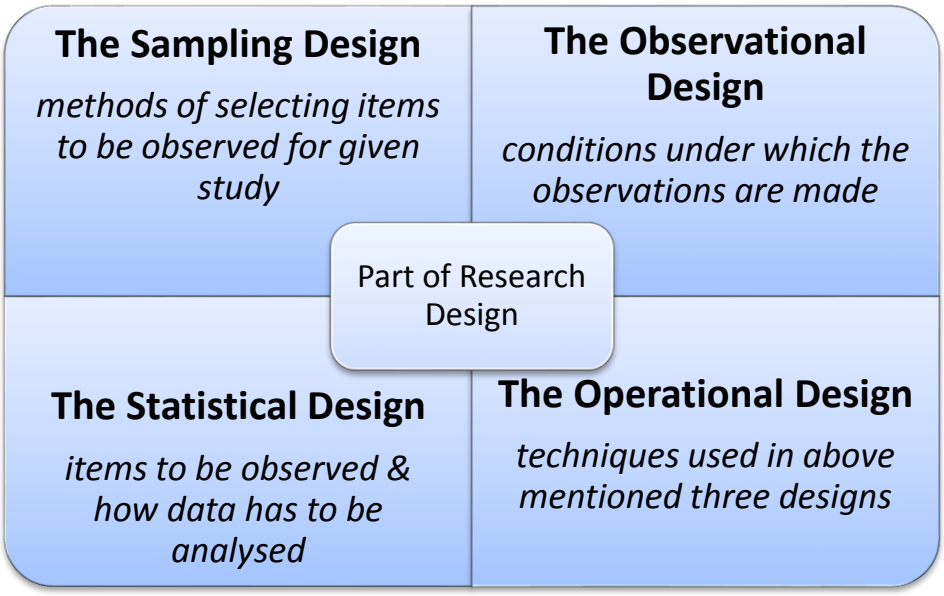
1. The definition of the key words, technical terms/and phrases used in the statement of the problem. Continuing the example cited above, the key words will be 'Growth', 'Secondary sector'

2. Basic assumptions of the research

3. Criteria for the selection of the problem

Research Design

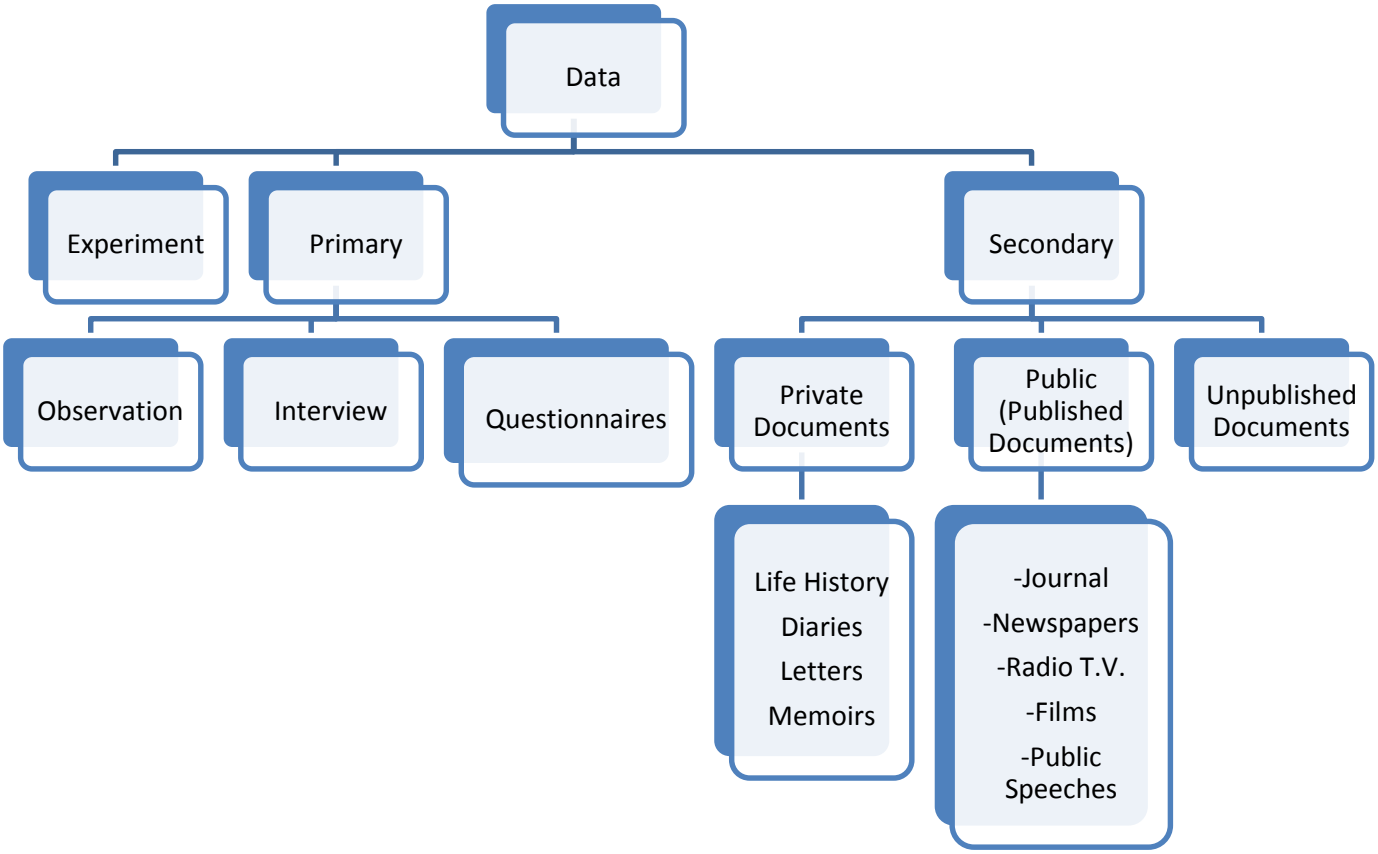
"A research design is the arrangement of conditions for collection and analysis of data in a manner that aims to combine relevance to the research purpose with economy in procedure" [1]



(explanation to follow)

[1] Claire Sellitz and others, *Research Methods in Social Sciences*,1962,Page 50

Data Collection



Experiment: includes researcher's own parameters, there is no rigid process. The researcher according to the research problem controls factors and assess the effect on the set up/ the situation.

Primary Data: Primary source of data is the most dependable and accurate relative to Secondary Data.

The method is time consuming and costly, however compensated by the reliability of it. A researcher must prepare questionnaires and interview questions with caution keeping in mind the view point of the person filling the questionnaire or giving the interview:

-easy to answer

-not too many questions

-questions should not be repetitive

-should be easy to understand

The **Observation Method** of primary source is based on one's own investigation minus the interview. The researcher observes the happenings and the situation is unaffected by the past or future behavior. It is not suitable in cases where large samples are required.

The **Secondary Data:** A researcher must wisely choose the secondary data keeping in mind the objective for the data can be inadequate or unsuitable or even both.

Evaluation of Secondary Sources of Data:

i. *What was the purpose of the study? Why was the information collected?*
ii. *Who was responsible for collecting the information? What qualifications resources, and potential bases are represented in the conduct of the study?*
iii. *What information was actually collected? How were units and concepts defined? How Direct were the measures used? How complete was the information?*
iv. *When was the information collected? Is the information still current or have events made the information obsolete? Were there specific events occurring at the time the data were collected that may have produced the particular results obtained?*
v. *How consistent is the information from one source with information from the other sources?*

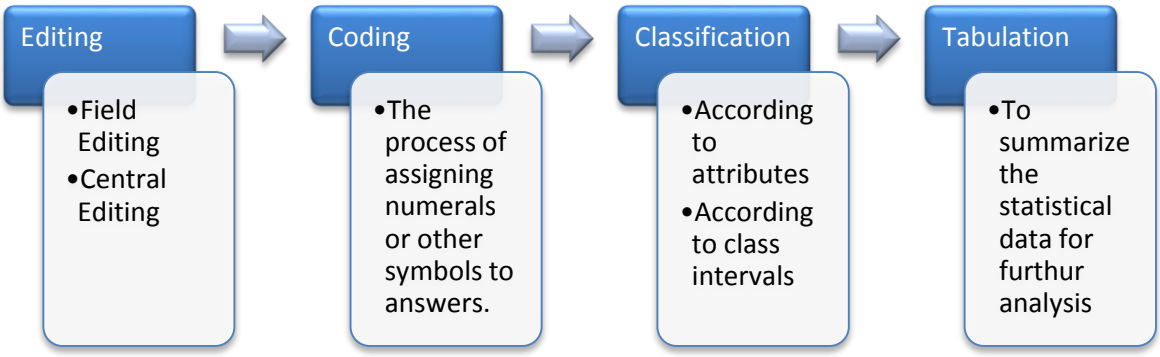
Processing and Analysis of Data

This is where the meaning of research starts materializing. Under this step, the data goes under editing, omitting, coding and classification followed by statistical tools being applied on data sieved so far.

Processing = editing + coding + classification + tabulation

Analysis = computation of data

Processing



1. Editing

A careful scrutiny of the data collected in order to discard what all is irrelevant to the objective of research.

Editing is done to assure that the data are accurate, consistent with other facts gathered, uniformly entered as completed as possible and have been well arranged to facilitate coding and tabulation.[2]

On the basis of stages or points at which editing is done, editing is of two kinds:

1. Field Editing

Editing the information collected by the medium of interviewing. It requires to putting the respondents' responses in an orderly manner. Ideally this sort of editing should be done on the very day of interview.

2. Central Editing

This type of editing implies, thorough checking of all the forms by the team of researchers, putting the data in its correct place. For instance, errors such as entry in the wrong column, entry recorded in months when it should have been recorded in weeks, so forth.

2. Coding

Coding decisions should be usually taken at the designing stage of questionnaire.

[2] *Research and methodology methods and techniques,* revised second edition by C.R.Kothari

It is the process of assigning numerals or other symbols to answers so that responses can be put into a limited number of category or classes.

Such classes should be **exhaustive** (for each data item there must be a class) and **mutually exhaustive** (specific answer can be placed in only one cell).

3. Classification:

It is arranging the data into homogenous groups or classes on the basis of some common characteristics.

a. Classification according to attributes:

The data is divided on the basis of attributes. Ideally, one class is of those items having the attributes and the other class not having the specified attributes in the former class.

Whenever data are classified according to attributes, the researcher should make sure that the attributes' defined properly so that there is minimum ambiguity when comes to classifying the data.

b. Classification according to Class-intervals

Data relating to production, age, income, weight, so forth: quantitative in nature are classified according to class intervals.

For instance, you have a data of monthly income of household of a locality. Income ranging between Rs.10,000 to Rs.1,00,000. Then class intervals of range 10,000 can be made.

Rs.10,000-Rs.20,000

Rs.20,000-Rs.30,000… Rs.90,000-Rs. 1,00,000 where in each class interval, small is value is the lower class limit (LCL) and the larger value is called Upper Class Limit (UCL).

Range of a class limit= UCL – LCL, in this case it is 10,000.

Mid-Point of CI= (UCL+LCL)/2

i. *how many classes should be there? What should be their magnitude?*

There is no perfect answer to this. Typically, there can be 5-15 classes. The magnitude of class intervals can be both equal and unequal. Some statistians even adopt a formula to determine the size of class interval given by H.A.Surges

$$i = R/(1 + 3.3 \log N)$$

i= size of the class interval

R= range (here range is the difference between the largest and the smallest value presnt in the entire data)

N= Number of items to be grouped

Ii. *How to choose class limits?*

The researcher must take into consideration the criterion that the mid-point of a class interval and the actual average of items of that class interval should remain close ro each other as possible. The class limits should be located at multiples of 2, 5,10,100,etc.

Exclusive CI

10-20 (Read as 10 and **under** 20)

20-30 (read as 20 and **under** 30)

30-40… (read as 30 and **under** 40)

Eg: if there is figure 30 it will fall under the CI 30-40 and **NOT** 20-30.

Exclusive CI should be used when the data is continuous in nature,c apable of being measured In fractions as well.

Inclusive CI

11-20 (including 20)

21-30 (including 30)

31-40… (Including 40)

Eg: if there is a figure 30, it will fall under the CI 21-30.

Inclusive CI should be used when data is discrete, measured and stated only in integers.

iii. *How to determine the frequency of each class?*

There are two ways,

1. Mechanical aid: In case of large inquiries and surveys, there are sorting machines, which can be either hand operated or run in electricity. The very fact that having access to such kind of machinery is expensive; this method is most widely used by large firms for complicated research.
2. Tally sheets: The most frequently used method.
   For each item a vertical stroke is marked against the class group it falls in. After every four vertical strokes, the fifth line for the item falling in the same group is indicated as a horizontal line through the said four vertical lines representing the total numerical value of five.
   An illustration is as follows:

Table: 01, Tally Mark

Income Groups (in Rs)	Tally Mark	Number of Families (Class frequency)
10,000-20,000	~~IIII~~ III	8
21,000-30,000	~~IIII~~	5
31,000-40,000	III	3
Total		16

4. Tabulation

Tabulation is information set out in tabular form.[3] It is to put the data into a more presentable, systematic and concise manner in order to do further calculations on the data collected.

Depending upon the magnitude of data, one can make the data manually or use computers/or statistical software. In case of small inquiries, use the former method and otherwise the former one.

Classification of Tables:

i. Simple: information about one or more groups of independent questions. It generally results in one way tables which answers questions about one characteristic of data only

ii. Complex: depicts division of data in two or more categories and is designed to give information concerning one or more sets if inter-related questions. Results in two way tables, three way tables or still higher order tables, called Manifold Tables which gives information of many interrelated variables. Such tables are also called cross tabulation.

Guidelines for Tabulation

- Table should have an apt title and should be placed above the body of the table
- Each table should be given a distinct number to facilitate easy reference
- The units of measurement under heading should be clearly specified
- Explanatory footnotes, if any, concerning the table should be placed directly beneath the table, along with the reference symbols used in the table.
- Sources of the data used in the table should be clearly specified
- Those columns, of which data has to compared, should be placed side by side.
- Ditto marks should not be used in the tables
- Decimal points, and the (-) and (+) signs should be in perfect alignment
- Miscellaneous and exceptional items, if any, should be usually placed in the last row of the table

[3] wordnetweb.princeton.edu/perl/webwn

Problems While Processing

1. 'Don't Know'
   If the group of 'DK' response is small, then it was of little relevance.
   However, if they constitute a majority part of responses, then it can affect the outcome of the data collected.
   Remedy:
   a. Formulate better questions
   b. While interviewing , explain the questions so that it is easier to answer for the interviewer
   c. Make DK responses as a separate category in tabulation

2. Use or percentages
   Cautions while using percentages:
   1. Two or more percentages must not be averaged unless each is weighted by the group size from which it has been derived.
   2. Percentage decreases can never exceed 100 per cent and as such for calculating the percentage of decrease, the higher figure should invariably be taken as the base.

Analysis

Detailed Examination of the elements or structure of something is referred to as Analysis. In terms of research, it is searching for patters of relationship that exist among the data groups. [4]

In case of experiments, analysis involves, estimating population parameters and testing hypothesis for further interpretation of the data.

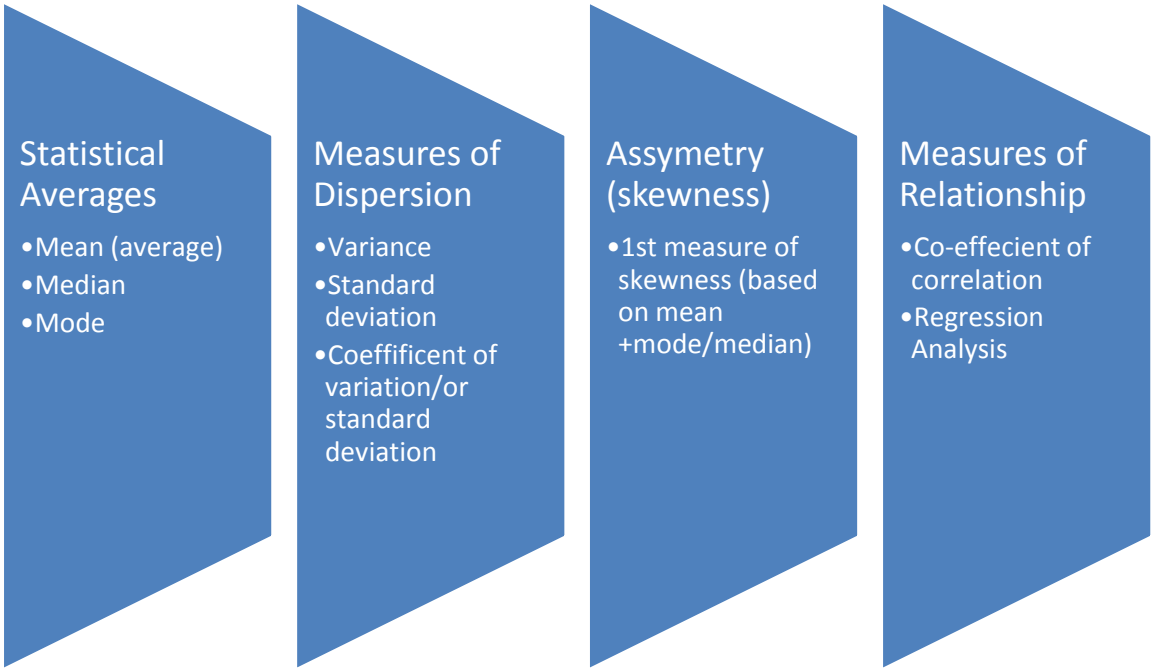
(**types of analysis)**

[4] *Research and methodology methods and techniques,* revised second edition by C.R.Kothari

Statistics in Research

Statistics is the scientific tool, on application of which we can conclude our research, test our hypothesis and base interpretations.

It is the tool which facilitates (**more intro)**



Measures of Central Tendency

(Mean, Median, Mode)

These measures indicate about the items that have a tendency to cluster. Such an indicator is considered as the most representative figure for the entire data.

$$\text{Mean } (\bar{x}) = \frac{\sum \text{Xi}}{\text{n}} = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n}$$

$\bar{x}$ = Symbol for mean

$\sum$ = Symbol for summation

X_i = Value of the i^{th} term X, I = 1,2,…,n

n= total number of items

In case of a frequency distribution,

$\bar{X} = \frac{\sum f_i X_i}{\Sigma f_i} = \frac{f_1 X_1 + f_2 X_2 + \cdots + fnXn}{f_1 + f_2 + \cdots fn = n}$

Sometimes, instead of simple mean, we might calculate weighted mean, for a ''realistic average''

$$\bar{X}_w = \frac{\Sigma_{w_i} X_i}{\Sigma w_i}$$

$\bar{X}_w = weighted\ item$
W_i= weight of the i^{th} item X
X_i= value of the i^{th} item X

Median

Median (M)= Value of $\left[\frac{n+1}{2}\right]^{th}\ item$

Mode

Measures of Dispersion

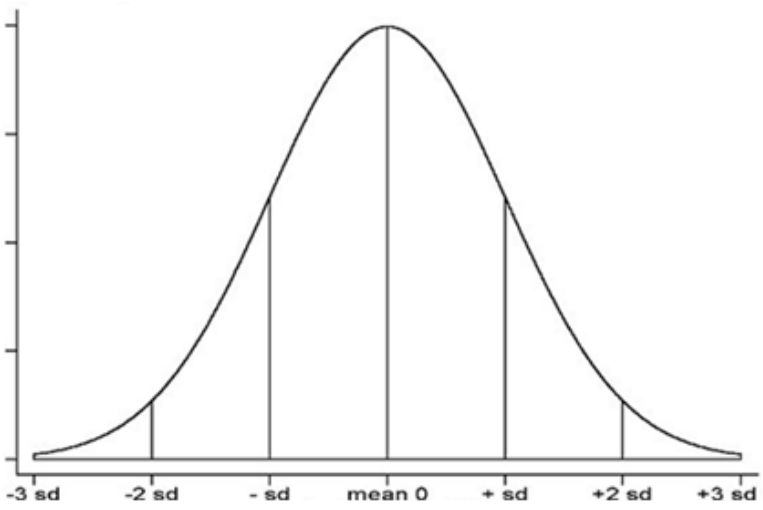
In order to measure the scatter of the data, such measures of dispersion are calculated.

It gives an idea that, as to what extent the data deviates (standard deviation) from its central value and over what extent is it spread (range).

Range (R)	simplest measure of dispersion, it is the difference between the two extreme values of a series R= (Highest Value) – (Lowest value)
Mean Deviation (Averages of deviation of the value of an item from another average {mean/median/ mode})	Mean Deviation from mean $\delta_{\bar{X}} = \frac{\sum \lvert X_i - \bar{X} \rvert}{n}$, $\lvert X_i - \bar{X} \rvert$ are deviations from arithmetic average. Mean Deviation from Median $\boldsymbol{\delta_m} = \frac{\sum \lvert \boldsymbol{X_i - M} \rvert}{\boldsymbol{n}}$, $\lvert X_i -$ Mean Deviation from Mode $\delta_z = \frac{\Sigma \lvert X_i - Z \rvert}{n}$, $\lvert X_i - Z \rvert$ are deviation from Mode.
Standard Deviation (most widely used, denoted by '' (sigma)	

Measures of Asymmetry (Skewness)

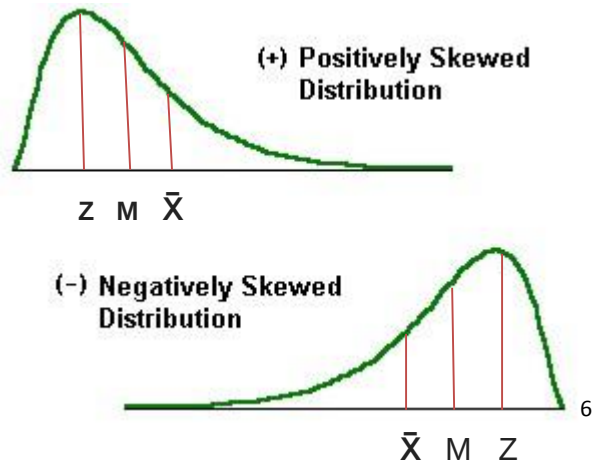
Perfectly Symmetrical (no skewness)



[5]

- $\bar{X}$ = M = Z (Mean = Mode = Median)
- Normal Curve, for Normal distributions
- Because the data is perfectly distributed on both sides, there is said to be No Skewness.

Positive and Negative Skewness



[6]

Sknewness shows the manner in which the items are clustered around the average.

The amount by which the balance exceeds on one side measures the skewness of the series.

Positive Skewness: Z<M< $\bar{X}$

Negative Skewness: $\bar{X}$<M<Z

Skewness implies the shape of the curve of the series, when the items of a given series are plotted on a graph.

[5] http://www2.connectseward.org/shs/da1/review07/Describingdistributions.htm
[6] http://grants.hhp.coe.uh.edu/doconnor/PEP6305/Topic%20002%20Organizing%20Data2.3.htm

Skewness = $\bar{X}$- Z

(Coefficient) j = $\frac{\bar{x}-Z}{\sigma}$

In Case Z is not well defined,

Skewness = 3($\bar{X}$-M)

(Coefficient) j = $\frac{3(\bar{x}-M)}{\sigma}$

Measures of Relationship

As a researcher, we might be interested in knowing how income and consumption are related or income and savings are related for a defined population, in that case we used the given below, measures of relationship.

If there are two variables, it is called a bivariate population otherwise, a multivariate population. For instance, if for every measurement of a variable X, we have a Y it is Bivariate in nature and if there was a Z series, it would be multivariate.

There are various methods for determining such a relationship between variables, but no method can tell us for certain that a correlation is indicative of casual relationship.[7]

Karl Pearson's Co-efficient of Correlation(r) / Simple Correlation

This is the most extensively used method for measuring the degree of relationship b/w **two** variables.

Assumptions of this method: [8]

1. There exists linear relationship between the two variables
2. The two variables are casually related which means that one of the variables is independent and the other one is dependent
3. A large number of independent causes are operating in both variables so as to produce a normal distribution

Karl Pearson's coefficient of correlation, r

(formulae)

Pointers:

- **'r'** is also called **product moment correlation** coefficient.
- Value of **'r'** lies between ±1

[7] *Research and methodology methods and techniques,* revised second edition by C.R.Kothari (Pg 138)

[8] *Research and methodology methods and techniques,* revised second edition by C.R.Kothari (Pg 139) One can check the book, for further statistical tools to measure relationship given qualitative variables and multivariate population.

- Positive value of 'r' indicates positive correlation between the two variables (i.e. the changes in both variables take place in the **same** direction)
- Negative value of 'r' indicates negative correlation existing in the bivariate population (i.e. the changes in both the variables take place in the **opposite** direction)
- 'r' = zero, there is **no** correlation between the variables.
- 'r' = +1, perfect positive correlation , 'r' = -1 perfect negative correlation
- The value of 'r' nearer to +1 or -1 indicates high degree of correlation between the two variables.

Simple Regression Analysis

Regression Analysis is used to make quantitative predictions of one variable from the values of another.[9]

A simple regression has only two variables, that is, a bivariate population is taken under consideration. One variable is defined as **Independent** while another is defined as **Dependent variable.** The basic relationship is between X and Y is given by:

$$\hat{y} = a + bX$$

$\hat{y}$ (read as: y hat), denotes the estimated value of Y, for a given value of X i.e. **Y** is a **dependent variable**, **X** is an **independent variable**.

Such an equation is called Regression Equation of **Y on X.**

Interpret: Each unit change on X produces a change of 'b' in Y, which is positive for direct and negative for inverse. 'b' is the slope of the given regression equation which measure the per unit change in Y.

'a' is the intercept term, which may/may not have an economic interpretation. It is that value of Y, when X = 0. For instance, if Y is consumption and X is income, then, if X=0, $\hat{y} = a$, implying that when income is zero there is consumption which is hypothetical in nature, hence, 'a' may or not have any interpretation except that it the intercept term while plotting the graph of the equation.

(give formulae)

Interpretation and Report Writing

Interpretation implies establishing inferences and conclusions from the statistical tool/ and experiments applied to the data collected.

In one sense, interpretation is concerned with relationships within the collected data, partially overlapping analysis. Interpretation also extends beyond the data of the study to include the results of other research, theory and hypotheses.[10]

[9] http://wordnetweb.princeton.edu/perl/webwn?s=regression%20analysis
[10] C.William Emory, *Business Research Methods,* p. 336.

Techniques of interpretation

Steps of Writing Report

Preparation of the final outline

- The final outline, is the blue print of the report. One ensures that all the material information is logically organized. An outline is the reminder of the points to be stressed upon in the report.

Preparation of the rough draft

- The draft should include, a brief of as to how the research was conducted: why the topic was chosen, the methods and procedure adopted to data, the statistical tools implemented, errors (if any), assumptions (in any).

Rewriting and polishing of the riough draft

- The rough draft is thoroughly examined to check if there is any information left out, if the report is has any gramattical mistakes and accordin edit/and omit.

Preparation of the final bibliography

- Listing down all the sources from where the data/ information was collected.
- Bibliography should be alphabetically arranged
- Example (*for books and pamphlets) :* Name of the author (*last name first),* Title (*in italics)*, Place, publisher and date of publication and number of volumes.
- Example (*for magazines and newspapers):* Name of author (*last name first)*, "Title of the Article", Name of the periodical (*in italics),* the volume, the date of the issue, the pagination.

Writing the final draft

- As a researcher concluded wiritng the report, he/she needs to ensure that the report is written in simple language enthuses people and create their interest in the research done.
- Words from the technical jargon should be avoided.
- Example from common experiences are recommended to be added
- Every report should be an attempt to solve an intellectual problem and its quality is judged by the ability of a reserach tp contribute to the solution of the problem or add another relevant prospective to the scenario.

Layout of the Research Report

A. Preliminary Pages

- Title and date
- Acknowledgement (Preface/Foreword)
- *Table of Contents*
- *List of Tables* and *illustrations.*

B. Main Text

(Each main section of the report must start on a new page)