Methodology

Chapter Objectives

- Research Methodology
- Data, Classifications and Tabulation
- Methodology of Study
- Tools and Techniques
6. Methodology

6.1 Research Methodology

The methodology used in the project is the core of the study and that should be first familiarized. This project is done in marketing thus it is necessary to define marketing first. "Marketing is the process by which matching of product with market is done and by this ownership transfer is done". This is the definition of marketing with which we can say that marketing is too very necessary for institutional sales also. Further, it will help us to define research because research is the heart of the study. "Research comprises of defining and redefining problems; formulating hypothesis or suggest solution; collecting, organizing and reaching conclusion, and at last carefully testing the conclusion; and at last carefully testing the conclusion whether they fit the formulating hypothesis". These are the meaning of marketing and research, but combining marketing and research, we get marketing research and it means." Marketing research is systematic gathering, recording, and analysis of data about problem relating to the marketing of goods and services". Here marketing research is defined as the research done to solve the problems of arises in the process of marketing.

The marketing research process is carried out according to a designated series of steps that are required to be taken in a chronological order. The major steps are:-

1. Problem Identification
2. Research design
3. Collection of Data
4. Data Analysis
5. Report Presentation

1. Problem Identification: The first step in marketing research process is to identify the problem chosen for investigation. The step is very
significant; if wrong problem is identified then the research result may be completely useless.

2. Preparing the Research Design:

A research design serves as a bridge between what has been established (the research objectives) and what is to be done, in the conduct of the study, to realize those objectives. It anticipates what the client will need in terms of results and the analytical work on the gathered data that will convert it to useful findings.

Preparation of such design facilitates research to be as efficient as possible yielding maximal information. The function of research design is to provide for the collection of relevant evidence with maximal effort, time and money. These can be achieved by keeping in mind the research purpose.

Research purpose may be grouped into four categories:

a) Exploration
b) Description
c) Diagnosis
d) Experimentation

The suitable design keeping in mind the purpose will minimizes the bias and maximizes the reliability of collection and analyzed.

Major types of designs

The specific designs are numerous and placing them under just three broader groups. These are

1. Descriptive
2. Experimental and
3. Quasi experimental.

Descriptive designs have that name because they describe phenomena without establishing association between factors. The data may
be (1) the behavioral variables of people (or the subjects) that are under study and (2) the situational variable that existed or are forthcoming.

Experimental designs are intended to demonstrate cause and effect relationships, as earlier mentioned. If they are measure concomitant variation, the degree of change in one variable, $y$ when the other variable, $x$ is changed, that is of greater benefit.

Quasi experimental designs as their name implies, lack the rigorous qualities of the real experiments, particularly a valid demonstration of variables association. Still, they are not descriptive ones either, because they do yield some quantitative indicators of association between their variables. Five types of quasi-experimental designs are there.

In addition to their frequent use as conclusive studies, descriptive designs are used for preliminary and exploratory studies.

There is a common denominator with which various design are placed in those three groups and that is their ability to measure causal association. Only conclusive research distinctly seeks to determine whether there are causal connection between an action that a decision maker is considering and the objective being sought. Bear in mind that absolute proof of causation cannot be provided by any research. One is that a suspected factor (hypothesis) that ostensibly varies with the sought objective does so by happenstance or as the result of some other factor, extraneous to the test. Such limitations, however, are inherent in any research design and do not depreciate the great worth of a causal design to a decision maker.

Conditions for Causal Relationships

Three conditions are necessary to justify making cause and effect statements:

1. Strong evidence must be obtained that an association exists between an action and an observed outcome.
2. There must be evidence that the action preceded the result (or outcome)
3. It must be evident that there are no other valid explanations for the observed results.

An approach to eliminating other possible causal factors is using statistical techniques that allow the researcher to measure and statistically remove their effects.

Descriptive designs: These cover research designs that are intended to produce accurate descriptions of variables relevant to the decision being faced, without demonstrating that some relationship exists between variables. Decision makers very often choose or a rewilling to accept descriptive data, which would permit only inferences to be drawn about causation. The obstacles to valid experimental studies may be so severe that they must be rules out. In the present study two types: (1) cross sectional, which comprises most of them, and (2) focus group was adopted.

Cross sectional designs produce a picture of the phenomena in which the decision maker is interested. Ad the data would be responses from (or observations of) a sample containing a large number of sources; it is a “cross section” of the situation. For many marketing decisions, cross sectional data are appropriate. Designs of cross sectional studies may be informal or implicit rather than rigorously described. When explicitly described, they would be simpler than designs of experimental studies, for they need to include only the nature and source of the data, the nature of expected results, and the analytical method. Apart from advantages in costs and time (compared with experimental methods) they are the appropriate choice for preliminary studies (to discern and define problems) and for exploratory studies (seeking optional actions) that is why it is adopted during the research study.

Focus group study

The designs of a focus group study are informal which is adopted in this study. These conversations are led and guided by a moderator, who has a standard way of initiating the discussion, of establishing a relaxed atmosphere, and of launching the conversation with the desired focus. The
objectives are definite, but aside from that, there is little structure, and the interviews are exploratory for whatever useful information turns up. The moderator must be well disciplined and prepared to steer the discussion onto pertinent subjects and to stimulate responses from everyone in the group. Focus groups are most useful in preliminary studies, as they inform a marketing manager of consumers' problems and needs often revealing unsuspected opportunities as well as complaints. They may be beneficial also in findings ideas for marketing campaigns, product improvement, and other marketing actions. They are so relatively easy to conduct and so intriguing that they are misused for conclusive purposes. It is said that clients too often fail to use them to understand consumers, but instead interpret them as showing how consumers behave.

3. Collection of data

Collection of data means the method that are to be employed for obtaining the require information for units under investigation. The method of collection of data depends upon the nature, objective and scope of investigation. It is also depends the availability of money and time.

The two types of data are:

a) Primary Data

b) Secondary Data

Primary Data: These are those data which are obtain for the first time by investigation from himself, in other words the data originally obtain in the presence of investigation are known as Primary Data.

Secondary Data: These data are those data, which have already being collected by some other person for their purpose that is the data taken from publish or unpublished records like news paper, periodicals, magazines etc.

The collection of data is the part of field work, which is to be carried during training. Here we can say once the researcher finalize the problem and
research design step, he has to conduct actual data collection operation and this is fieldwork.

4. Data Analysis

In order to extract meaningful information from the data collected, the data analysis is carried out. The data are first edited, coded and then encoded for the purpose of analyzing them.

5. Report Presentation

After the collected data is analyzed and interpreted, the job of the marketing researcher is to present results of the research in the form of the systematically typed or printed research.

![Diagram for the Process on working on Project](image-url)

Fig 6] Diagram for the Process on working on Project
6.1.1. SOME EXAMPLES OF DIFFERENT TYPES OF RESEARCH

- **Analysis**: classes of data are collected and studies conducted to discern patterns and formulate principles that might guide future action.

- **Case study**: the background, development, current conditions and environmental interactions of one or more individuals, groups, communities, businesses or institutions is observed, recorded and analyzed for stages of patterns in relation to internal and external influences.

- **Comparison**: two or more existing situations are studied to determine their similarities and differences.

- **Correlation-prediction**: statistically significant correlation coefficients between and among a number of factors are sought and interpreted.

- **Evaluation**: research to determine whether a program or project followed the prescribed procedures and achieved the stated outcomes.

- **Design-demonstration**: new systems or programs are constructed, tested and evaluated experiment: one or more variables are manipulated and the results analyzed.

- **Survey-questionnaire**: behaviors, beliefs and observations of specific groups are identified, reported and interpreted.

- **Status**: a representative or selected sample of one or more phenomena is examined to determine its special characteristics.

- **Theory construction**: an attempt to find or describe principles that explain how things work the way they do.

- **Trend analysis**: predicting or forecasting the future direction of events.
6.2 Data, Classification and Tabulation.

For the further analysis, which is to be carried out collection of data, is done. This is a very necessary step for the interpretations and analysis, which is the essence of the project work. For this sampling is to be done.

1. **Purpose of sampling**: - Sampling is the only tool which help to know the characteristics of the population by examining only a small part of its. The law of statistical regularity says, “a moderately large number of its items chosen at random from a large group are almost sure on the average to possess the characteristics of large group.

   **Sampling** is that part of statistical practice concerned with the selection of individual observations intended to yield some knowledge about a population of concern, especially for the purposes of statistical inference. In particular, results from probability theory and statistical theory are employed to guide practice.

The sampling process consists of five stages:

- Definition of population of concern .
- Specification of a sampling frame, a set of items or events that it is possible to measure.
- Specification of sampling method for selecting items or events from the frame.
- Sampling and data collecting.
- Review of sampling process.

**SAMPLING TYPES**

**Simple sampling**

In this case, all elements of the frame are treated equally and it is not subdivided or partitioned. One of the sampling methods below is applied to the whole frame.
Stratified sampling
Where the population embraces a number of distinct categories, the frame can be organized by these categories into separate strata or demographics. One of the sampling methods below is then applied to each stratum separately, maintaining the same balance in numbers as exists in the population and resulting in an improvement in precision.

Cluster sampling
Where items in the population are clustered, sampling can reflect this to minimize costs. For example, in a national survey by personal interview, many people will be remotely located and costly to reach. Cluster sampling locates the frame in areas of concentrated habitation.

Multistage sampling
Sampling method
Within any of the types of frame identified above, a variety of sampling methods can be employed, individually or in combination.

Random sampling
In Random sampling, every combination of items from the frame, or stratum, has an equal probability of occurring. It guarantees that the sample is representative of the frame but is infeasible in many practical situations. It is a type of probability sampling.

Systematic sampling
Selecting (say) every tenth name from the telephone directory is simple to implement and is an example of systematic sampling. Though simple to implement, asymmetries and biases in the structure of the data can lead to bias in results. It is a type of no probability sampling.

Convenience sampling
Sometimes called, grab sampling, this is the method of choosing items arbitrarily and in an unstructured manner from the frame. Though almost
impossible to treat rigorously, it is the method most commonly employed in many practical situations.

**Sample size**

Where the frame and population are identical, statistical theory yields exact recommendations on sample size. However, where it is not straightforward to define a frame representative of the population, it is more important to understand the cause system of which the populations are outcomes and to ensure that all sources of variation are embraced in the frame. Large numbers of observations are of no value if major sources of variation are neglected in the study.

**Sampling and data collection**

Good data collection involves:

- Following the defined sampling process
- Keeping the data in time order
- Noting comments and other contextual events
- Recording non-responses

**Review of sampling process**

After sampling, a review should be held of the exact process followed in sampling, rather than that intended, in order to study any effects that any divergences might have on subsequent analysis. A particular problem is that of non-responses.

**Non-responses**

In survey sampling, many of the individuals identified as part of the sample may be unwilling to participate or impossible to contact. In this case, there is a risk of differences, between (say) the willing and unwilling, leading to bias in conclusions. This is often addressed by follow-up studies, which make a repeated attempt to contact the unresponsive and to characterize their similarities and differences with the rest of the frame.
**Systematic sampling** is the selection of every $k^{th}$ element from a sampling frame, where $k$, the sampling interval, is calculated as:

\[
k = \frac{\text{Number in population}}{\text{Number in sample}}
\]

Using this procedure each element in the population has a known and equal probability of selection. This makes systematic sampling functionally similar to simple random sampling. It is however, much more efficient and much less expensive to do.

The researcher must ensure that the chosen sampling interval does not hide a pattern. Any pattern would threaten randomness. A random starting point must also be selected.

**Stratified sampling** is a method of sampling from a population in statistics.

**Multistage sampling** (adopted in the research study) is a complex form of cluster sampling. Using all the sample elements in all the selected clusters may be prohibitively expensive or not necessary. Under these circumstances, multistage cluster sampling becomes useful. Instead of using all the elements contained in the selected clusters, the researcher randomly selects elements from each cluster. Constructing the clusters is the first stage. Deciding what elements within the cluster to use, is the second stage. The technique is used frequently when a complete list of all members of the population does not exist.

Examples of nonprobability sampling include:

**Convenience sampling** - members of the population are chosen based on their relative ease of access, which is adopted in the study. To sample friends, co-workers, or shoppers at a single mall, are all examples of convenience sampling.

**Snowball sampling** - The first respondent refers a friend. The friend also refers a friend, etc.

**Judgmental sampling or Purposive sampling** - The researcher chooses the sample based on who they think would be appropriate for the study. This is
used primarily when there is a limited number of people that have expertise in the area being researched.

Case study - The research is limited to one group, often with a similar characteristic or of small size.

ad hoc quotas - A quota is established (say 65% women) and researchers are free to choose any respondent they wish as long as the quota is met.

Even studies intended to be probability studies sometimes end up being non-probability studies due to unintentional or unplanned characteristics of the sampling method. In public opinion polling by private companies (or organizations unable to require response), the sample can be self-selected rather than random. This often introduces an important type of error: self-selection error. This error sometimes makes it unlikely that the sample will accurately represent the broader population. Volunteering for the sample may be determined by characteristics such as submissiveness or availability. The samples in such surveys should be treated as non-probability samples of the population, and the validity of the estimates of parameters based on them unknown.

1.1. Sample Design: - For doing this project, the sampling technique used for this is “Convenience sampling technique”

1.2 Size of the sample: - The appropriate sample for a study may be defined as that sample which fulfill the requirement of efficiency, representative ness, reliability, and flexibility that is, the sample must be small enough to avoid unnecessary expenses and large enough to avoid sample error beyond the limit of tolerance. The sample should yield the desired information with the required level of reliability at minimum cost.

1.3 Segmentation: - The region was district Durg, for the purpose of data collection; it has been divided into different profile as per population, urban and rural.

2. Method: - The study needs more accuracy as well as investigation so the best method is to collect primary data.
The survey for this project was done by "Direct personal interview". This method helps the investigator to personally come by asking questions and helping them to fill questionnaire. Thus information is obtained is accurate. Data obtained by this method is reliable to a greater extent.

Secondly Personal observation method was adopted in which the investigator attempts to view the world through the eyes of the target group. This is done by systematically observing their behavior, by recalling past observations, or by reinterpreting previous experiences.

Thirdly the "key informant" method obtains information from community residents who are in positions to know the community well. This information gathering technique uses an intense and personal format for gathering information from community knowledgeable.

2.1 Source of Data: - The primary Data for the research is collected by visiting to the purchaser, personal interview of consumer and dealers, and focus group study and the structured questionnaire has been used for collecting these data. The secondary data for the research is collected thru journals, magazines, and newspapers and through Internet.

6.3 Methodology of study
Type of research:
The research study has a qualitative approach. It has the some characteristics of exploratory and of descriptive one with evaluation of causal relationship.
Research Approach: The research study has the qualitative and quantitative approach.

The thesis, deals with effect of marketing communication process on and purchase of consumer durable goods by a consumer/buyer/customer, is based on survey/schedule and focus group study carried during the research period. In this research study, the consumer survey was scheduled as festival (August to December) and off festival (February to July) season canvassed in approximate 550 households and 12 focus group selected from urban and

211
rural blocks spread over the geographical area of the District Durg (C.G.). The reference period was 180 days.

For each item of durable goods listed in the category group, information was collected on process, factors, effect incurred on the item separately for purchase and consumption, during reference periods. The schedule also collected some other household particulars including age, sex and educational level of each respondent.

**Durable goods:**

Household consumer goods other than food, fuel and clothing/footwear are, according to usual NSS practice, classified into durable goods and a residual “miscellaneous goods” category. Durable goods are distinguished from miscellaneous goods by having a longer expected lifetime of use (roughly, one year or more), though some petty durables such as pens, torches, spectacles and, as well as books, are by convention included in miscellaneous goods.

But in our research study the **durable goods** has been classified into **three category groups** as mentioned below:

- **GROUP: A - ENTERTAINMENT** - TV, DVD, Radio, Tape/Stereo, etc.
- **GROUP: B - HOUSEHOLD APPLIANCES**-washing machines, microwave oven, etc.
- **GROUP: C - UTILITIES**- vehicles as cars, motorcycles; mobiles, etc.

**Study Area:** District Durg selected urban and rural areas as per (SAMPLING)

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Composition</th>
</tr>
</thead>
<tbody>
<tr>
<td>I.</td>
<td>Place with population less than 10000, urban</td>
</tr>
<tr>
<td>II.</td>
<td>Place with population 10000 – 50000, urban</td>
</tr>
<tr>
<td>III.</td>
<td>Place with population 50000 and above, urban</td>
</tr>
<tr>
<td>IV.</td>
<td>Place with population 10000 - affluent areas</td>
</tr>
<tr>
<td>V.</td>
<td>Place with population 10000- 50000 other areas</td>
</tr>
<tr>
<td>VI.</td>
<td>Place with population less than 10000, rural</td>
</tr>
<tr>
<td>VII.</td>
<td>Place with population less than 50000, rural</td>
</tr>
<tr>
<td>VIII.</td>
<td>Place with population 50000 and above, rural</td>
</tr>
</tbody>
</table>

Exhibit 6.2. Sampling.
The data for the study was collected through a structured questionnaire. Five role (as defined by Philip Kotler, namely, Initiator, Influencer, decider, buyer and user) played by family members in purchase of a durable product were examined for the study. The questionnaire used for survey was divided into three parts. The first part was to identify the buyer wants and desire, the second part was focused on the nature of the buyer and third part of the questionnaire included the personal details and addressed to the family members who actually played that role in purchase of that product.

The Sample

A non-probability convenience sample was adopted. The sample consisted of 550 households who either owes, recently purchased or going to purchase at least one of the three durables categories. While selecting the sample items, care were taken to ensure that the families selected represented different income and age groups, professions, joint and nuclear structures, those with single earning members as well as double income families.

For the purpose of the survey, a family was defined as consisting of husband, grandparents and other member who did not fall in the above categories. The nomenclature of the family members depended upon the composition of the family. The age of member had no bearing on the name given to him/her and the members were designated as husband, wife, children, grandparents etc., depending upon the number of generations in that household.

Family Size

The majority of families surveyed had four members each, followed with three and five members each; in keeping with an independent finding that a family across urban and rural India consists of 3.9 individuals.

Monthly Household Income

The majority of household who purchased a durable product either fell in to the more than Rs.25000 (26.8%) or in Rs.15000-20000 (26.6%) income
bracket, followed by families with monthly household income of Rs 10000-15000 (17.8%) and Rs 20001-25000 (17.0%). Only 11.8% families had a monthly household income of up Rs. 10000.
6.4 Tools and Techniques

**ANALYSIS OF VARIANCE (ANOVA)**

Analysis of variance (abbreviated as ANOVA) is an extremely useful techniques concerning researches in the fields of economics, biology, education, Psychology, sociology, business/industry and in researches of several other disciplines. This technique is used when multiple sample cases are involved. As, the significance of the difference between the means of two samples can be judged through either z-test or the t-test, but the difficulty arises when we happen to examine the significance of the difference amongst more than two sample means at the same time. The ANOVA technique enables us to perform this simultaneous test and as such is considered to be an important tool of analysis in the hands of a researcher. Using this technique, one can draw inferences about whether the samples have drawn from population having the same mean.

The ANOVA technique is important in the context of all situations where we want to compare more than two populations. In such circumstances one generally does not want to consider all possible combinations of two populations at a time, which might require great number of test for decision-making. This would also consume lot of time and money, and even then certain relationships may be left unidentified (particularly the interaction effects). Therefore, one quite often utilized the ANOVA technique and through it investigates the differences among the means of all the populations simultaneously.

**WHAT IS ANOVA?**

Professor R.A. Fisher was the first man to use the term 'Variance' and in fact, it was he who developed a very elaborate theory concerning ANOVA, explaining its usefulness in practical field. Later on Professor Snedecor and many others contributed to the development of this technique. ANOVA is essentially a procedure for testing the difference
among different groups of data for homogeneity. The essence of ANOVA is that the total amount of variation in a set of data is broken down into two types, that amount which can be attributed to chance and that amount which can be attributed to specified causes." There may be variation between samples and also within sample items. ANOVA consists in splitting the variance for analytical purposes. Hence, it is a method of analyzing the variance to which a response is subject into its various components corresponding to various sources of variation.

Thus, through ANOVA technique one can, in general, investigate any number of factors, which are hypothesized or said to influence the dependent variable. One may as well investigate the differences amongst various categories within each of these factors, which may have a large number of possible values. If we take only one factor and investigate the differences amongst its various categories having numerous possible values, it is said to use one-way ANOVA. In a two or more way ANOVA, the interaction (i.e. inter-relation between two independent variables/factors), if any, between two independent variables affecting a dependent variable can as well be studied for better decisions.

**THE BASIC PRINCIPLE OF ANOVA**

The basic principle of ANOVA is to test for differences among the means of the populations by examining the amount of variation within each of these samples, relative to the amount of variation between, the samples. In terms of variation within the given population, it is assumed that the values of \((X_{ij})\) differ from the mean of this population only because of random effects i.e., there are influences on \((X_{ij})\) which are unexplainable, whereas in examining differences between populations we assume that the difference between the mean of the \(j\)th population and the grand mean is attributable to what is called a 'specific factor' or what is technically described as treatment effect. Thus while using ANOVA, we assume that each of the samples is
drawn from a normal population and that each of these populations has the same variance. We also assume that all factors other than the one or more being tested are effectively controlled. This, in other words, means that we assume the absence of many factors that might affect our conclusions concerning the factor(s) to be studied.

In short, we have to make two estimates of population variance viz., one based on between samples variance and the other based on within samples variance. Then the said two estimates of population variance are compared with F-test, wherein we work out:

\[
F = \frac{\text{Estimate of population variance based on between samples variance}}{\text{Estimate of population variance based on within samples variance}}
\]

This value of F is to be compared to the F-limit for given degrees of freedom.

If the F value we work out is equal or exceeds the F -limit value (to be seen from F tables), we may say that there are significant differences between the sample means.

**ANOVA TECHNIQUE**

One -way (or single factor) ANOVA: Under the one – way ANOVA, only one factor is considered and then observe that the reason for said factor to be important is that several possible types of samples can occur within that factor. It is then determined if there are differences within that factor. The techniques involves the following steps:

(i) Obtain the mean of each sample i.e. obtain

\[
\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots, \bar{X}_k
\]

When there are k samples.
(ii) Workout the mean of the sample means as follows:

\[
\bar{X} = \frac{\bar{X}_1 + \bar{X}_2 + \bar{X}_3 + \ldots + \bar{X}_k}{\text{No. of samples} \ (k)}
\]

(iii) Take the deviations of the sample means from the mean of the sample means and calculate the square of such deviations which may be multiplied by the number of items in the corresponding sample, and then obtain their total. This is known as the sum squares for variance between samples (or SS between). Symbolically, this can be written:

\[
SS \text{ between} = n_1 \left( \frac{\bar{X}_1 - \bar{X}}{\bar{X}_1 - \bar{X}} \right)^2 + n_2 \left( \frac{\bar{X}_2 - \bar{X}}{\bar{X}_2 - \bar{X}} \right)^2 + \ldots + n_k \left( \frac{\bar{X}_k - \bar{X}}{\bar{X}_k - \bar{X}} \right)^2
\]

(iv) Divide the result of the (iii) step by the degrees of freedom between the samples to obtain variance or mean square (MS) between samples. Symbolically, this can be written:

\[
MS \text{ between} = \frac{SS \text{ between}}{(k-1)}
\]

where \((k-1)\) represents degrees of freedom (d.f.) between samples.

(v) Obtain the deviations of the values of the sample items for all the samples from corresponding means of the samples and calculate the squares of such deviations and then obtain their total. This total is known as the sum of squares for variance within samples (or SS within).

\[
SS \text{ within} = \sum (X_{1i} - \bar{X}_1)^2 + \sum (X_{2i} - \bar{X}_2)^2 + \ldots + \sum (X_{ki} - \bar{X}_k)^2
\]

\[i=1,2,3,\ldots\]
(vi) Divide the result of (v) step by the degrees of freedom within samples to obtain the variance or mean square (MS) within samples. Symbolically, this can be written:

\[
SS \text{ within} \\
MS \text{ within} = \frac{SS \text{ within}}{(n-k)}
\]

where \((n-k)\) represents degrees of freedom within samples, \\
\(n\) = total number of items in all the samples i.e., \(n_1 + n_2 + ... + n_k\) \\
i = number of samples.

(vii) For a check, the sum of squares of deviations for total variance can also be worked out by adding the squares of deviations when the deviations for the individual items in all the samples have been taken from the mean of the sample means. Symbolically, this can be written:

\[
SS \text{ for total variance} = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (X_{ij} - \bar{X})^2
\]

This total should be equal to the total of the result of the (iii) and (v) steps explained above i.e.,

\[
SS \text{ for total variance} = SS \text{ between} + SS \text{ within}.
\]

The degrees of freedom for total variance will be equal to the number of items in all samples minus one i.e., \((n-1)\). The degrees of freedom for between and within must add up to the degrees of freedom for total variance i.e.

\[
(n-1) = (k-1) + (n-k)
\]

This fact explains the additive property of the ANOVA technique.
(viii) Finally, F-ratio may be worked as under:

\[
\frac{MS\ between}{F\ -\ ratio} = \frac{MS\ within}{MS\ within}
\]

This ratio is used to judge whether the difference among several sample means is significant or is just a matter of sampling fluctuation. For this purpose we look into the table, giving the values of F for given degrees of freedom at different levels of significance.

If the worked out value of F, as stated, is less than the table value of F, the difference is taken as insignificant i.e., due to chance and the null-hypothesis of no difference between sample means stands. In case the calculated value of F happens to be either equal or more than its table value, the difference is considered as significant (which means the samples could not have come from the same universe) and accordingly the conclusion may be drawn. The higher the calculated value of F is above the table value, the more definite and sure one can be about his conclusions.
SETTING UP ANALYSIS OF VARIANCE TABLE

For the sake of convenience the information obtained through various steps stated above can be put as under:

**Analysis of Variance Table for One-way ANOVA**

*(There are k samples having in all n items)*

<table>
<thead>
<tr>
<th>Source of variation</th>
<th>Sum of squares (SS)</th>
<th>Degrees of Freedom (d.f)</th>
<th>Mean Square (MS) (This is SS divided by d.f.) and is an estimation of variance to be used in F-ratio</th>
<th>F-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between samples or categories</td>
<td>$n_1 \left( \bar{X}_1 - \bar{X} \right)^2 + ...$</td>
<td>$(k-1)$</td>
<td>$\frac{SS \text{ between}}{(k-1)}$</td>
<td>$\text{MS between} \quad \text{and MS within}$</td>
</tr>
<tr>
<td></td>
<td>$+ n_k \left( \bar{X}_k - \bar{X} \right)^2$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$\sum (X_{ij} - \bar{X})^2$</td>
<td>$(n-k)$</td>
<td>$\frac{SS \text{ between}}{(n-k)}$</td>
<td></td>
</tr>
<tr>
<td>Within samples or categories</td>
<td>$i = 1, 2, 3, .......$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>$\sum (X_{ij} - \bar{X})^2$</td>
<td>$(n-1)$</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>$i = 1, 2, ..., j = 1, 2, ...$</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
TWO WAY ANOVA

Two-way ANOVA technique is used when the data are classified on the basis of two factors. For example, a business firm may have its sales data classified on the basis of different salesmen and also on the basis of sales in different regions. In a factory, the various units of a product produced during a certain period may be classified on the basis of different varieties of machines used and also on the basis of different grades of labour. Such a two-way design may have repeated measurements of each factor or may not have repeated values. The ANOVA technique is little different in case of repeated measurements where we also compute the interaction variation.

(a) ANOVA technique in context of two-way design when repeated values are not there:

As we not have repeated values, we cannot directly compute the sum of squares within samples as we had done in the case of one-way ANOVA. Therefore, we have to calculate this residual or error variation by subtraction, once we have calculated (just on the same lines as we did in the case of one-way ANOVA) the sum of squares for total variance and for variance between varieties of one treatment as also for variance between varieties of the other treatment.

The various steps involved are as follows:

(i) Use the coding device, if the same simplifies the task.

(ii) Take the total of the values of individual items (or their coded values as the case may be) in all the samples and call it T.

(iii) Work out the correction factor as under:

\[
\text{Correction factor} = \frac{(T)^2}{n}
\]

(iv) Find out the square of all the item values (or their coded values as the case may be) one by one and then take its total. Subtract the correction factor from this total to obtain the sum of squares of deviations.
for total variance or total SS

\[(T)^2\] = \[\sum X^2_{ij} \div n\]

(v) Take the total of different columns and then obtain the square of each column total and divide such squared values of each column by the number of items in the concerning column and take the total of the result thus obtained. Finally, subtract the correction columns or (SS between column).

(vi) Take the total of different rows and then obtain the square of each row total and divide such squared values of each row by the number of items in the corresponding row and take the total of the result thus obtained. Finally, subtract the correction factor this total to obtain the sum of squares of squares of deviations for variance between rows (or SS between rows).

(vii) Sum of squares of deviations for residual or error variance can be worked out by subtracting the result of the sum of (v) th steps from the result of (iv) th step stated above. In other words,

Total SS- (SS between column +SS between rows) = SS for residual or error variance.

(viii) Degrees of freedom (d.f.) can be worked out as under:

d.f. for total variance = (c.r-1)  
d.f. for variance between columns = ( c-1)  
d.f. for variance between rows = (r-1)  
d.f. for residual variance = (c-1)(r-1)  
Where c= number of columns  
r = number of rows

(ix) ANOVA table can be set up in the fashion as shown:
### Analysis of Variance Table for TWO-way ANOVA

*There are k samples having in all n items*

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>Sum of squares (SS)</th>
<th>Degrees of Freedom (d.f)</th>
<th>Mean Square (MS)</th>
<th>F-ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Between columns treatment</td>
<td>( \sum \frac{(T_i)^2}{n_i} - \frac{(T)^2}{n} )</td>
<td>(c-1)</td>
<td>SS between columns (\frac{(c-1)}{(c-1)})</td>
<td>MS between Columns</td>
</tr>
<tr>
<td>Between Rows treatment</td>
<td>( \sum \frac{(T)^2}{n} )</td>
<td>(r-1)</td>
<td>SS between Rows (\frac{(r-1)}{(r-1)})</td>
<td>MS residual</td>
</tr>
<tr>
<td>Residual or error</td>
<td>Total SS - (SS between columns + SS between rows)</td>
<td>(c-1) (r-1)</td>
<td>SS residual (\frac{(c-1)(r-1)}{(c-1)(r-1)})</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>( \sum \frac{X^2_{ij} - (T)^2}{n} )</td>
<td>(c,r-1)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In the table, \( c = \) number of columns
\( r = \) number of rows

\( SS \) residual = Total SS - (SS between columns + SS between rows).
Thus, \textit{MS} residual or the residual variance provides the basis for the \textit{F}-ratio concerning variation between columns treatment and between rows treatment. \textit{MS} residual is always due to the fluctuations of sampling, and hence serves as the basis for the significance test. Both the \textit{F}-ratio are compared with their corresponding table values, for given degrees of freedom at a specified level of significance, as usual and if it is found that the calculated \textit{F}-ratio concerning variation between columns equal to or greater than its table value, then the difference among columns means is considered significant. Similarly, the \textit{F}- ratio concerning variation between rows can be interpreted.

(b) ANOVA technique in context of two way when repeated values are there: For this measure we calculate the sum of squares, degrees of freedom and sum of squares for variance within samples as in one way ANOVA. We then find left over sum of squares and left over degrees of freedom, which are used for what is known as "interaction variation" (Interaction is the measure of inter relationship among the two different classifications). After making all these computations ANOVA table is set up for drawing inferences.

6.4.1 Standard deviation inference

The standard deviation measures the amount of variability in the distribution of a variable. Thus, the more that the individual data points differ from each other, the larger the standard deviation will be. Conversely, if there is a great deal of similarity between data points, the standard deviation will be quite small. The standard deviation describes the standard amount variables differ from the mean. Examining differences in variability could be useful for anticipating further analyses: in the study, Because equal variances are an assumption of many inferential statistics, this information is important to a data analyst.
Variance: a measure of how data points differ from the mean. Standard deviation is a measure of the difference from the mean. Large standard deviation means the data is spread widely around the mean; units are the same as the data itself.

Computer software package SPSS vers.12 is used for data tabulation, analytical and drawing statistical inferences in the research study. (Refer Chapter-7)

***************

REFERENCES:
6. Further References, refer on page 309, of this present Thesis.