Chapter - 3

Methodology and Database

This chapter deals a brief description of the study area and the techniques used for the selection of sample and processing of data under the following headings.

3.1. Description of the Study Area

3.2. Sampling Design

3.3. Collection of Data

3.4. Method of Analysis

3.1. Description of the Study Area

3.1.1. Location of the Study Area

The study was conducted in Bangalore Urban District of Karnataka, which is situated in the Southern part of the Karnataka state.

Karnataka has a rich heritage, inspiring its people to create a bright future. With its special geographical location full of variety—its rivers, hills, valleys, plains, forests and resources—the State is known for its tourist and industrial potential. Its long history of over 2,000 years has left many beautiful forts, tanks,
temples, mosques and towns of historical importance to the posterity. These old towns have grown to be industrial, commercial and educational centers. They are provided with all Modern facilities.

The State of Karnataka, confined roughly within 11°35' North and 18°30' North latitudes and 74°5' East and 78°35' East longitudes, is situated on a table land where the Western and Eastern Ghat ranges converge into the Nilgiri hill complex. It is located in the Western part of the Deccan Peninsular region of India. The State is bounded by Maharashtra and Goa States in the North and North-West; by the Arabian Sea in the West; by Kerala and Tamilnadu States in the South and by the State of Andhra Pradesh in the East. The State extends, to about 750 km from North to South and about 400 km from East to West.

**Physiography**

Physiographically of Karnataka State forms part of two well defined macro-regions of Indian Union; the Deccan Plateau and the Coastal plains and Islands. The State has four physiographic regions as follows:

1. **Northern Karnataka Plateau**: Northern Karnataka Plateau comprises of the districts of Belgaum, Bidar, Bijapur, Bagalkot and Gulbarga.

2. **Central Karnataka Plateau**: Central Karnataka Plateau covers the districts of Bellary, Chikmagalur, Chitradurga, Davanagere, Dharwad, Gadag, Haveri Raichur, Koppal and Shimoga.
3. **Southern Karnataka Plateau**: The Southern Karnataka Plateau covers the districts of Bangalore, Bangalore Rural, Hassan, Kodagu, Kolar, Mandya,

4. **Karnataka Coastal Region**: The Karnataka Coastal Region, which extends between the Western Ghats edge of the Karnataka Plateau in the east and the Arabian Sea in the west, covers Dakshina Kannada, Udupi and Uttara Kannada districts.

**Climate**

The State enjoys three main types of climates. For meteorological reasons, the State has been divided into three sub-divisions namely (a) Coastal Karnataka (Dakshina Kannada, Udupi and Uttara Kannada districts), (b) North Interior Karnataka (Belgaum, Bidar, Bijapur, Bagalkot, Dharwad, Gadag, Haveri, Gulbarga, Raichur and Koppal districts) and (c) South Interior Karnataka (the remaining districts of Bangalore Rural, Bangalore, Bellary, Chikmagalur, Chitradurga, Davanagere, Kodagu, Hassan, Kolar, Mysore, Chamarajnagar, Mandya, Shimoga and Tumkur districts). The Tropical Monsoon climate covers the entire coastal belt and adjoining areas. The climate in this region is hot with excessive rainfall during the monsoon season i.e., June to September. The Southern half of the State experiences hot, seasonally dry tropical savanna climate while most of the northern half experiences hot, semi-arid, tropical steppe type of climate.
India's pride, Bengaluru is nearly 500 years old and has grown from a small
time settlement when Kempe Gowda, the architect of Bengaluru, built a mud fort
in 1537 and his son marked the city boundaries by erecting four watch towers.
Today Bengaluru has grown well beyond those four towers into a sprawling
metropolis of more than 6 million people and is referred to as the Silicon Valley of
India - accounting for more than 35 percent of India's software exports.
Bengaluru's temperate climate, high quality educational, scientific and technology
institutions coupled with a thriving IT and Bio-Technology and manufacturing
industry makes Bengaluru one of the most sought after global destinations.

Bangalore city was purposively selected for the study Bangalore city the
capital of Karnataka State is a cosmopolitan area with a diverse population cutting
across different socio economic and cultural groups which gives it the rich
diversity to undertake consumer studies. Bangalore exemplifies every aspect of the
Indian milieu with economic progress the demand for consumer durables increase
for a variety of reasons which gives rise to a retail Revolution. This is the trend
that has effectively captured in Bangalore which can then be generalized for the
whole country
Map 2. Map of Karnataka
Map 3. Map of Bangalore (Study Area)
Table 1. Geographical Classification of Bangalore City

<table>
<thead>
<tr>
<th>Sl. No.</th>
<th>Zone Names</th>
<th>Number of wards</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Yelahanka</td>
<td>11</td>
</tr>
<tr>
<td>2</td>
<td>Mahadevapura</td>
<td>17</td>
</tr>
<tr>
<td>3</td>
<td>Dasarahalli</td>
<td>8</td>
</tr>
<tr>
<td>4</td>
<td>Rajarajeshwari</td>
<td>21</td>
</tr>
<tr>
<td>5</td>
<td>West</td>
<td>35</td>
</tr>
<tr>
<td>6</td>
<td>South</td>
<td>46</td>
</tr>
<tr>
<td>7</td>
<td>East</td>
<td>44</td>
</tr>
<tr>
<td>8</td>
<td>Bommanahalli</td>
<td>16</td>
</tr>
</tbody>
</table>

Source: BBMP Website

3.2. Sampling Design

3.2.1. Selection of Sample Size

After turning into greater Bangalore the Bangalore city consists of 27 assembly constituencies, other than this the city are divided into division and subdivision and then into 196 wards at the area of 740 sq km. and the average population per ward was 40,277, the highest being 47,000 and lowest 30,000 and I selected random samples from these wards.

3.2.2. Selection of the Sample Respondents

The Cluster sampling method was followed and sample respondents were post stratified based on their income, education and age. In Bangalore city out of
196 wards made 10 clusters, data was collected from the samples randomly drawn from each cluster. Totally 500 sample respondents were selected from all these wards.

3.3. Collection of Data

For evaluating the specific objectives of the necessary study, primary data was obtained from the selected respondents through Personal interview method with the help of pre-tested and structured schedule. Respondents were questioned in Kannada and English.

The data collected from the respondents related to their Age, Education and Income and also costs incurred on the purchase of consumer durables and attributes related to the consumer durables and then brand’s perception on each consumer durables.

In addition to primary data collected from sample respondents, secondary data was also obtained. Secondary data was collected on sales and purchase of each consumer durables for the source on internet.

3.4. Method of Analysis

Tabular analysis was used for estimating the income of the respondents. Various Econometric techniques were used like Cluster Analysis, Multidimensional Scaling and Logistic Regression and even basic statistics techniques.
Cluster Analysis: This is one of the multivariate techniques used to classify a sample of entities, individual or objects into smaller number of mutually exclusive groups based on the similarities among the entities.

Logistic Regression: Logistic regression is a specified form of regression that is formulated to predict and explain a binary (two-group) categorical variable which is used to categorize into buyers or non-buyers.

Multidimensional Scaling: This technique was used to understand consumer perception with regard to product or brand and for perceptual mapping. The objective is to transform consumer judgments of preference/similarity to brand into distances represented in multidimensional space.

3.4.1. Logit Analysis

The logit model is used in capturing the qualitative responses in the dependent variable. In the present study, it is employed to know the effects of age, education, income and family size on consumer durables purchase. When the dependent variable is dichotomous in nature, application of linear regression model leads to erroneous results. Under such circumstances, binary-choice models are used which assume that individuals are faced with a choice between two alternatives and that the choice they make depends on the characteristics of the individuals. The purpose of these models is to determine the probability that an individual with a given set of attributes would choose one or the other alternative.
The simplest form of the model involves the dependent variable assuming a binary response, which takes values of 1 and 0. The commonly used qualitative response model in Economic analysis is the Linear Probability Model, the Logit Model and Probit Model.

**Linear Probability Model (LPM)**

The regression form of the model is,

\[ Y = \alpha + \beta x + u_i \quad (1) \]

It is estimated through Ordinary Least Squares (OLS) method which undergoes some disadvantages which are as follows:

1) The Variance of the disturbance \((u_i)\) will not be homoscedastic \((E(u_i)=0)\) which is against one of the assumptions of OLS.

2) The assumption of normality is no longer tenable for LPM- because like \(y\), \(u_i\) takes only two values

3) Estimated probabilities lie outside the range of 0 and 1.

Further, the estimation of probability of OLS assumes that the probability increases linearly with the explanatory variable, i.e., the marginal or incremental effect of the explanatory variable remains constant throughout, which will not happen in reality.
To overcome these discrepancies, the Logit and the Probit models are preferred. These models are developed based on the logistic cumulative distribution function and normal cumulative distribution function, respectively.

In the present study, the logit model is preferred to the probit model owing to the computational ease.

The logit model based on the logistic probability is specified as

$$ P_i = F(z_i) = F(\alpha + \beta_i X_i) = \frac{1}{1 + e^{-z_i}} \quad (2) $$

where,

$$ z_i = \alpha + \beta_i X_i $$

After simplifying the above formula for estimation purpose, one can write the logit model as

$$ z_i = \ln \left( \frac{P_i}{1 - P_i} \right) = \alpha + \beta_i X_i = L_i \quad (3) $$

$P_i$ = Probability of age,

$1 - P_i$ = Probability of age, education and income

$\beta_i$ = Coefficient to be measured

$X_i$ = Independent variable

$E$ = Base of the natural logarithms, which is approximately equal to 2.72.

$L_i$ is called the logit as it follows logistic regression.
Pi/(1-Pi) is the odds ratio in favour of age, education and income on Consumer durables

Given the limitations of OLS, the maximum likelihood techniques was used in estimating the logit co-efficient. One model was fitted to each variable. The marginal effect of the ith variable on Pj is given by the first derivative of P with respect to Xi.

\[
\frac{dp}{dxi} = \beta_i(1-Pi)
\]

Thus the elasticity of this probability is

\[
E_{pi} = \beta_i (1-Pi)X_i
\]

The independent variables considered in the model are described below.

1) X1 Average Education in the family
2) X2 Average Age of the family
3) X3 Average Income of the family
4) X4 Average Family size

3.4.2. Multidimensional Scaling

Multidimensional scaling (MDS) is also known as perceptual mapping, a procedure that allows a researcher to determine the perceived relative image of a set of objects (firms, products, ideas, or other items associated with commonly held perceptions). The purpose of MDS is to transform consumer judgments of
overall similarity or preference (e.g., preference for stores or brands) into distances represented in multidimensional space.

**Analytical methods for producing perceptual maps**

Factor analysis of customer ratings.

- Use focus groups and/or prior research to identify important attributes for product class.
- Ask 100 to 300 customers to rate the products with which they are familiar on a set of metric measurement scales for these attributes.
- Determine number of dimensions.
- Use factor loadings to name underlying dimensions.
- Plot average factor scores (across customers) for each brand in map.
- Using a metric measure of stated preferences, regress preference on the factor scores to get the slope of the preference vector.
- If multiple segments are desired, a finite mixture regression model can be used to produce multiple preference vectors.
- Factor analysis approach depends on the identification of a well-specified set of attributes.
- Not model-based – involves a two-stage process.
Maps can also be formed by fitting models to preference/choice data:

- A persons-by-stimuli two-way data matrix.
- Attributes may or may not be utilized.
- These models allow for statistical inferences to be made, unlike the factor analysis approach and MDS.
- Preference vectors represent segments rather than individuals, easing interpretation.

Maps can be formed from cross-tabulated data using correspondence analysis.

- Using Multidimensional Scaling (MDS)
  - Ask customers to rate the similarity of pairs of brands on a metric scale – no attributes are involved!
  - These measures are averaged across all customers (or segments of customers) to produce a proximity matrix whose entries represent the similarity/dissimilarity among the products.
  - Given the number of dimensions, MDS program finds locations of brands that minimize “stress,” a measure of lack of fit.
  - Interpreting dimensions will be more difficult without attributes:
    - Include some attribute labels (“Prestige”) along with brands and plot the labels alongside the brands.
Measure attribute ratings as in factor analysis, regress attribute ratings on brand locations, and use regression coefficients (like loadings) to name dimensions.

- Comparison of techniques:
  - Ratings on attributes can be linked to potential product improvements.
  - When factor analysis is used, we cannot identify any perceptual dimensions that are not represented by the attribute measures.
  - Similarity measures indicate which products are likely to be considered as substitutes.
  - Similarity scaling is useful if some attributes of a product are difficult to scale (allure of a perfume) or difficult to articulate (ambiance of a restaurant).
  - Difficult to name and interpret similarity-based dimensions.

- Aggregate vs. Disaggregate analysis:
  - MDS can be performed at individual level, segment level, or market level.
  - Combining information across individual-level maps is a challenge.
  - For a more aggregated map, average similarity judgments prior to analysis.
  - Clustering the coordinates from individual-level maps is not recommended.
  - Mixture of both (INDSCAL):
Form perceptual map on basis of all consumers (a shared space); brands are represented as points.

- Represent respondents’ preference weights for the dimensions separately; each respondent is represented by a vector whose slope is determined by regression coefficients.

- Length of vector represents % variance explained.

- Space can get very crowded, difficult to interpret.

- Segment-level analysis alleviates crowding.

• Metric vs. Nonmetric MDS
  o Rank orders of pairs used for nonmetric MDS (which pair is most similar? Least similar?)
  o All output is metric, meaning that the following do not change the solution:
    - Rotation about the origin.
    - Reflection of axes.
    - Entire solution can be stretched or compressed.
  o Metric data contains more information.

• Vector vs. Ideal Point (unfolding) Preference Representations:
  o Vector models assume more (or less) of a dimension is better (economy, performance, prestige).
  o Preference ordering of brands is found by dropping perpendicular lines from the brands to the consumer’s vector.
Ideal point models assume some intermediate level of a dimension is preferred (sweetness).

Brands closest (in a Euclidean distance sense) to a consumer’s ideal point are more preferred.

Ideal point may be found by having respondent evaluate his/her ideal brand along with the others, or ideal points may be estimated for each respondent using specialized estimation procedures.

- **Internal vs. External Analysis:**
  
  - Internal analyses estimate the brand coordinates.
  
  - In external analyses, the coordinates of brands are assumed given or known, for example, from previous analyses.

  - With internal analyses, brand coordinates can be reparameterized using attribute information to aid in the interpretation of the derived dimensions.

  - Better to do all analyses in one step rather than multiple steps!

**How to do an MDS?**

The researcher performs three basic steps to do a multidimensional scaling analysis:

- Gathers measures of similarity or preference across the entire set of objects to be analyzed.
• Uses MDS techniques to estimate the relative position of each object in multidimensional space.
• Identifies and interprets the axes of the dimensional space in terms of perceptual and/or objective attributes.

Objectives of MDS
• To identify latent dimensions (primary needs) affecting consumer behavior.
• To obtain comparative evaluations of objects when the specific bases of comparison are unknown or undefined.

Perceptual mapping and MDS in particular, is most appropriate for achieving two objectives:

1) As an exploratory technique to identify unrecognized dimensions affecting behavior.
2) As a means of obtaining comparative evaluations of objects when the specific bases of comparison are unknown or undefined.

Similarities vs. Preferences Data?

Both bases of comparison can be used to develop perceptual maps, but with differing interpretations:

• Similarity-based Perceptual maps – represent attribute similarities and perceptual dimensions of comparison but do not reflect any direct insight into the determinants of choice.
• Preference-based Perceptual maps – reflect preferred choices but may not correspond in any way to the similarity-based positions, because respondents may base their choices on entirely different dimensions or criteria from those on which they base comparisons.

• Either can be used in MDS, though interpretation will differ (how?).

• Common to use similarities in MDS, overlay preference vectors using regression analysis.

Research Design of MDS

Identification of all relevant objects to be evaluated:

• Errors of omission, inclusion of irrelevant objects.

• More than 4x as many objects as dimensions – results in many comparisons for 2+ dimensions.

• Overfitting can be a big problem if too few objects are used.

• Perceptual mapping techniques can be classified into one of two types based on the nature of the responses obtained from the individuals concerning the object:

• Decompositional method – measures only the overall impression or evaluation of an object and then attempts to derive spatial positions in multidimensional space that reflect these perceptions. This technique is typically associated with MDS.
• Compositional method – an alternative approach that uses several multivariate techniques in forming an impression or evaluation based on a combination of specific attributes.

Research Design of MDS

Perceptual maps can be generated through decompositional or compositional approaches:

• Decompositional approaches are the “traditional” and most common MDS method requiring only overall comparisons of similarity between objects.

• Compositional approaches are used when the research objectives involve comparing objects on a defined set of attributes.

• The number of objects to be evaluated is a trade-off between:
  - A small number of objects that facilitate the respondents’ task.
  - Four times as many objects as dimensions desired (i.e., 5 objects for one dimension, 9 objects for two dimensions…) to obtain a stable solution.

Deriving and Validating a MDS Solution

• Stress measures (lower values are better) represent a MDS solution’s fit.

• Researchers can identify a degenerate MDS solution which is generally Problematic by looking for:
  - A circular pattern of objects suggesting that all objects are equally similar, or
A multi-clustered solution in which objects are grouped at two ends of a single continuum.

• The appropriate number of dimensions for a perceptual map is based on:
  o A subjective judgment as to whether a solution with a given dimensionality is reasonable.
  o Use of a scree plot to identify the ‘elbow’ where there is a substantial improvement in fit.
  o Use of $R^2$ as an index of fit – measures of .6 or higher are considered acceptable.

• External analysis, such as is performed by PREFMAP, is considered preferable in generating ideal points relative to internal analysis. The most direct validation method is a split-sample approach. Multiple solutions are generated by either splitting the original sample or collecting new data. Validity is indicated when the multiple solutions match.

Assumptions of MDS Analysis

• Assumptions = none, other than the correct type of data is collected for the procedure used (metric vs. nonmetric MDS).

• Multidimensional scaling has no restraining assumptions on the methodology, type of data, or form of the relationships among the variables. But there are three perception requirements:
1. *Variation in Dimensionality* - respondents may vary in the dimensionality they use to form their perceptions of an object (although it is thought that most people judge in terms of a limited number of characteristics or dimensions).

2. *Variation in Importance* - respondents need not attach the same level of importance to a dimension, even if all respondents perceive this dimension.

3. *Variation Over time* - judgments of a stimulus in terms of either dimensions or levels of importance are likely to change over time.

**Deriving the MDS Solution and Assessing Overall Fit**

- Determining an object's position in the perceptual map.
- Selecting the dimensionality of the perceptual map using a stress measure (lower stress means better fit).
- Fit will improve when the number of dimensions increases.
- Index of fit, like $R^2$, can be used.
- Perceptions or Similarities Data?
- Determining an Object's Position in the Perceptual Map.
- Dimensionality of the Perceptual Map.
Interpreting the MDS Results

- Properly fitting.
- Subjective Procedures – may be best for affective, highly intangible, or emotional dimensions.
- Once the Perceptual map is obtained, the two approaches – compositional and decompositional – again diverge in their interpretation of the results. The differences in interpretation are based on the amount of information directly provided in the analysis (e.g., the attributes incorporated in the compositional analysis vs. their absence in the decompositional analysis) and the generalizability of the results to the actual decision-making process.

Compositional vs. Decompositional Methods

- For *compositional methods*, the perceptual map can be directly interpreted with the attributes incorporated in the analysis. The solution, however, must be validated against other measures of perception, because the positions are totally defined by the attributes specified by the researcher.
- For *Decompositional methods*, the most important issue is the description of the perceptual dimensions and their correspondence to attributes.

Validating the Results

- Validation efforts are problematic.
- In comparing two maps, if the positions vary, the researcher cannot specify whether the objects are viewed differently, the perceptual dimensions vary, or both.
Any MDS solution must deal with two specific issues which complicate efforts to validate the results:

- The only output of MDS which can be used for comparative purposes involves the relative positions of the objects. Thus, although the positions can be compared, the underlying dimensions have no basis for comparison. If the positions vary, the researcher cannot determine whether the objects are viewed differently, the perceptual dimensions vary, or both.

- Systematic methods of comparison have not been developed and integrated into the statistical programs. The researcher is left to improvise with procedures that may could general but not specific concerns.

**Approaches to Validating the MDS Results**

- Split-Sample Analysis.

- Comparison of Decompositional vs. Compositional Solutions.

### 3.4.3. Cluster Analysis

Cluster analysis is the name for a group of multivariate techniques whose primary purpose is to group objects based on the characteristics they possess. Cluster analysis classifies objects (e.g., respondents, products, or other entities) so that each object remains very similar to others in the cluster with respect to some predetermined selection criterion. The resulting clusters of objects should then exhibit high internal (within-cluster) homogeneity and high external (between-cluster) heterogeneity. Thus, if the classification is successful, the objects within
clusters will be close together when plotted geometrically, and different clusters will be far apart.

In cluster analysis, the concept of the variate is again a central issue, but in a quite different way from other multivariate techniques. The cluster variate is the set of variables representing the characteristics used to compare objects in the cluster analysis. Because the cluster variate includes only the variables used to compare objects, it determines the “character” of the objects. Cluster analysis is the only multivariate techniques that does not estimate the variate empirically but instead uses the variate as specified by the researcher. The focus of cluster analysis is on the comparison of objects based on the variate, not on the estimation of the variate itself. This makes the researcher’s definition of the variate a critical step in cluster analysis.

Cluster analysis has been referred as Q analysis, typology construction, classification analysis, and numerical taxonomy. This variety of names is due in part to the usage of clustering methods in such diverse disciplines as psychology, biology, sociology, economics, engineering, and business. Although the names differ across disciplines, the methods all have a common dimension: classification according to natural relationships [1,2,3,6,12,16]. This common dimension represents the essence of all clustering approaches. As such, the primary value of cluster analysis lies in the classification of data, as suggested by “natural” groupings of the data themselves. Cluster analysis is comparable to factor analysis.
in its objective of assessing structure. But cluster analysis differs from factor analysis in that cluster analysis groups objects, whereas factor analysis is primarily concerned with grouping variables.

Cluster analysis is a useful data analysis tool in many different situations. For example, a researcher who has collected data by means of a questionnaire may be faced with a large number of observations that are meaningless unless classified into manageable groups. Cluster analysis can perform this data reduction procedure objectively by reducing the information from an entire population or sample to information about specific, smaller subgroups. For example, if we can understand the attitudes of a population by identifying the major groups within the population, then we have reduced that data for the entire population into profiles of a number of groups. In this fashion, the researcher has a more concise, understandable description of the observations, with minimal loss of information.

Objectives of Cluster Analysis

1. **Taxonomy description.** The most traditional use of cluster analysis has been for exploratory purposes and the formation of a taxonomy—an empirically based classification of objects. As described earlier, cluster analysis has been used in a wide range of applications for its partitioning ability. But cluster analysis can also generate hypothesis related to the structure of the objects. Yet, although viewed principally as an exploratory
technique, cluster analysis can be used for confirmatory purposes. If a proposed structure can be defined for a set of objects, cluster analysis can be applied, and a proposed typology (theoretically based classification) can be compared to that derived from the cluster analysis.

2. **Data Simplification.** In the course of deriving a taxonomy, cluster analysis also achieves a simplified perspective on the observations. With a defined structure, the observations can be grouped for further analysis. Whereas factor analysis attempts to provide "dimensions" or structure to variables, cluster analysis performs the same task for observations. Thus, instead of viewing all of the observations as unique, they can be viewed as members of a cluster and profiled by its general characteristics.

3. **Relationship identification.** With the clusters defined and the underlying structure of the data represented in the clusters, the researcher has a means of revealing relationships among the observations that was perhaps not possible with the individual observations. Whether analysis such as discriminate analysis are used to empirically identify relationships, or the groups are subjected to more qualitative methods, the simplified structure from cluster analysis many times portrays relationships or similarities and differences not previously revealed.