CHAPTER IV
RESEARCH METHODOLOGY

Introduction

Research methodology plays an important role in any research. It includes research design, data collection, analysis and interpretation of results. The term ‘methodology’ comprises this whole process. The final results of a research depend on the methodology that we are employing and methodology depends on the type of data needed to answer the research questions. Social science researchers use either quantitative research methods, qualitative research methods or both (triangulation). But there is a controversial argument between social science researchers and scientific researchers regarding the use of these two methods in the social sciences and debates on quantitative and qualitative research methods is still continuing. However, these two approaches help to understand the socio-economic realities of the society. The qualitative approach uses non-numeric data and the quantitative approach uses numeric data. Selecting an appropriate methodology for specific research depends on the research objectives and research questions which are to be answered through the research. The research methods must lead to comprehensive and clear results at the end of the research.

The questionnaire is a measuring tool (Oppenheim, 1992) loosely; a questionnaire consists of a series of questions, checklists, attitude scales and a variety of other approaches in a structured sequence. They are used to provide descriptive and or analytical information which is suitable for statistical analysis.

Questionnaires usually involve large samples (upwards of 300) and are costly so it is essential to plan the research approach. Who to question, types of questions to ask, sample size, inherent biases, and these are among the factors that affect questionnaire measurement, specification and procedures (Oppenheim, 1992). Questionnaires need exploratory work, design and planning before any specification can be established. A certain rigidity of questioning and sampling procedure is needed to maintain the statistical validity which makes them relatively inflexible. Questionnaire assessments of, say, consumption of fuel wood, fodder, and food
grains are, for example, indirect unless a weighting measurement is included and are dependent on the accuracy of recall by the respondent (Oppenheim, 1992).

The methodology used in the study is an integrated methodology, where traditional schedule based data collection and processing is integrated with the modern, statistical as well as qualitative analysis. The former complements the latter. The methodology which follows the traditions of social science research (Kundu, 1992; Wood, 1996) and the latest developments in socio-economic research have the following components:

1. Field survey (primary data).
2. Collection of documented data (secondary data).
4. Analysis and interpretation of women’s data (of SHGs and Non-SHGs).

In selecting the most appropriate tool, the following considerations were useful: the Uses, Resources, Familiarity, Significance and the categories of people (women) involved. There are several ways of collecting the appropriate data which differs considerably in the context of money, cost, time and other resources at the disposal of the researcher. For the present study, both primary and secondary data have been collected and used for analysis.

Further, in research, there are various points of departure a researcher can choose from. In positivist approach, the researcher assumes that there is a truth to be discovered and that reality is value free, a-historical and cross-cultural. Science should, therefore, be neutral or value free. Further, a careful distinction between scientifically established objective meanings and subjective meanings are made.

The purpose of this chapter is to: discuss our research philosophy in relation to other philosophies; expound our research strategy, including the research methodologies adopted; and introduce research instruments that we have developed and utilized in the pursuit of our research.
Research Philosophy

A research philosophy is a belief about the way in which data about a phenomenon should be gathered, analyzed and interpreted. The term *epistemology* (what is known to be true) as opposed to *doxology* (what is believed to be true) encompasses the various philosophies of research approach. The purpose of science, then, is the process of transforming things believed in into things known: *doxa to episteme*. Two major research philosophies have been identified in the Western tradition of science, namely, the *positivist* (sometimes called scientific) and *interpretivist* (also known as anti-positivist) (Galliers, 1991).

Methodological Issues and the Rationale

While time constraints did not allow the use of the longitudinal approach, which is often the most preferred in such studies, the methodology, simple as it may sound, provides the opportunity to establish a baseline for a future longitudinal assessment. Mitchell (1989; 1990; and 1991) concluded that research on institutional arrangements for resource management has focused on *ex-post* studies of specific resource management programmes and projects with an emphasis on descriptive as opposed to predictive approaches. The present study, following this tradition, takes on a descriptive as opposed to predictive approach. Similarly, in the process and outcomes analysis, the case study has been the dominant research design. By concentrating on real world case studies, not only the researcher can test the applicability of some framework and model and tools, but she can assess such methods within a relevant context, which should enhance the replicability of the methodology and the results.

Hence, the study has used a post-positivist approach with less emphasis on aspects of modelling, benefit-cost analysis and analytical (statistical) approaches. This post-positivist approach is not biased towards quantification, but addresses adequately issues of uncertainty, values and socio-historical and behavioural contexts. Mitchell (1989: 20) concluded that phenomenological approaches to resource management have proved useful. He characterised phenomenological approaches as those that concentrated on the understanding of the human/environment interface by focusing on human attitudes, experiences and actions. The
choice of a post-positivist methodology is not to mean that the positivist approaches are not valid. They are very valid. In fact, the post-positivist approaches are valid inasmuch as they contribute to the body of theory, of course.

Sjoberg and Nett (1968) highlighted the strengths and weaknesses of the structured and un-structured interview format. They also concluded that structured interviews provide a means to standardise responses, facilitate the verification of theories and hypotheses, and provide greater reliability than unstructured interviews. They also noted that structured interviews can introduce bias, as researchers may impose their own categories and may have a tendency to oversimplify reality.

A Review of Research Methodologies

Qualitative Research (Interpretive)

*Interpretivists* contend that only through the subjective interpretation of, and intervention in, reality can that reality be fully understood. The study of phenomena in their natural environment is key to the *interpretivist* philosophy, together with the acknowledgement that scientists cannot avoid affecting those phenomena they study. They admit that there may be many interpretations of reality, but maintain that these interpretations are in themselves a part of the scientific knowledge they are pursuing. *Interpretivism* has a tradition that is no less glorious than that of positivism, nor is it shorter.

Quantitative Research (Positivistic, Statistical):

Quantitative research methods are research methods dealing with numbers and anything that is measurable. Quantitative research methodology calls for what is known as hard data in the form of numbers. Quantitative research methodology is about the collection of data in their numerical form. So they can be easily measured or counted. This research methodology is highly preferred by the positivist researchers who want to observe the social reality in terms of quantification and objectivity. By quantitative methods, researchers have come to mean the techniques of randomized experiments, paper and pencil “objective” test, multivariate statistical analysis, sample survey and the like (Cook and Reichardt, 1979; Hemmasi, 1994; 1996; Neuman, 2000). In social sciences, quantitative research methods express
different social phenomena in numbers. In the present case of research, the scholar has to use numerical data which are associated with household economic development in the context of self-help groups of women and outside of them. In the study, the researcher’s primary aim is to collect primary and secondary data. These data help to discover the past states, present states as well as future trends of the area by using available data. This is an additional advantage of using quantitative data.

Also the quantitative research methods help to generalize the existing social phenomena by testing samples. In this research, a sample survey for gathering primary data by using two different questionnaires and interviews have been made using the women of Nandivaram-Guduvancheri town panchayat in Kanchipuram district of Tamil Nadu. The women interviewed belong to self-help groups and also those who are currently not members of such SHGs. It has been useful to assess and understand how women of the town perceive household economic development in general, and evaluate their socio-economics about a number of aspects of livelihoods and work in suburban / peri-urban Chennai in particular so that some genuine analysis could be done to look at the household economic development in the town as seen through the women of the town.

Further the researcher has used SPSS (Statistical Package for Social Sciences) computer software to organize and analyze the data which are collected through the field survey. It is easy to handle quantitative data with SPSS programme and graphing, tabulating, and describing datasets. When making graphs using numerical data, it can be understood by anyone who does not have even a rudimentary knowledge of statistics. However, in the present research, quantitative data are important to explain the nature, value and knowledge of household economic development among the people of Nandivaram-Guduvancheri township and the conditions living and working there as well as in the state of Tamil Nadu, using the town of Nandivaram-Guduvancheri of Kanchipuram district as a case in point.

However, there are some limitations the researcher could identify in quantitative methods. Quantitative methods cannot clearly explain human feelings and thoughts like, for example, women’s household economic development
experiences. It is difficult to convert such things into numerical data. To overcome this problem, the researcher has used scaling for measuring certain contexts related items in the questionnaire. It helps respondents to have freedom to select a scale provided by the researcher in the tool.

Also dealing with a huge quantity of numbers sometimes makes for mistakes when handling them. Another limitation of quantitative research methods is that sometimes we cannot get the exact answers. However, quantitative data are important to our research in many ways and help to make general evaluation, regarding policy changes in household economic development related management, for example, impacting upon the women’s involvement in livelihood activities and work in the study area.

**Positivism**

Positivists believe that reality is stable and can be observed and described from an objective viewpoint (Levin, 1988); that is, without interfering with the phenomena being studied. They contend that phenomena should be isolated and that observations should be repeatable. This often involves manipulation of reality with variations in only a single independent variable so as to identify regularities in, and to form relationships between, some of the constituent elements of the social world.

Predictions can be made on the basis of the previously observed and explained realities and their inter-relationships. *Positivism has a long and rich historical tradition. It is so embedded in our society that knowledge claims not grounded in positivist thought are simply dismissed as unscientific and therefore invalid* (Hirschheim, 1985: 33). This view is indirectly supported by Alavi and Carlson (1992) who, in a review of 902 Information Science research articles, have found that all empirical studies are positivist in approach. Positivism has also had a particularly successful association with the physical and natural sciences. Some of the social science researches are no exception.

There has, however, been much debate on the issue of whether or not this *positivist paradigm* is entirely suitable for the social sciences (Hirschheim, 1985),
many authors calling for a more pluralistic attitude towards research methodologies (see for example, Kuhn, 1970; Bjørn-Andersen, 1985; Remenyi and Williams, 1996). While we would not elaborate on this debate further, it is germane to our study.

Indeed, some of the difficulties experienced in academic research such as the apparent inconsistency of results, may be attributed to the inappropriateness of the positivist paradigm for the domain. Likewise, some variables, or constituent parts of reality, might have been previously thought un-measurable under the positivist paradigm - and so went un-researched (Galliers, 1991).

Research Design of the Study

The present research has depended on both the primary sources of data and secondary sources of data.

Primary Sources of Data

The primary sources are the women of Nandivaram-Guduvancheri township. Three hundred women, with 150 each of women of the self-help groups operative in the township and those who do not belong to the self-help groups, have been randomly selected as the sample of the study so that a comparison between the household economic development of the two groups of women could be studied and understood.

Sample and Sampling

A sample is some part of a larger body specially selected to represent the whole. Sampling is the process by which samples for study are chosen. Sampling is taking any portion of a population or universe as a representative of that entire population or universe. For a sample to be useful, it should reflect the similarities and differences found in the total group. The main objective of drawing a sample is to make inferences about the larger population from the smaller sample. A census is a survey in which information is gathered from or about all members of a
population. For the present study Simple Random Sampling Method (SRS) was used for collection of information from the women of Nandivaram-Guduvancheri township by using the two custom-designed questionnaires, one for women of self-help groups and the other for women who do not belong to any SHGs. A sample of 300, 150 each to SHGs and non-SHG in the township chosen for the study.

The samples chosen have been interviewed using a custom-designed questionnaires, each with a distinct number of questions in each of its sections (Appendices 4.1 and 4.2). The samples have been chosen through a snowball sampling procedure, in which the scholar has chosen a select number of women from the town through the SHG functionaries and through them the rest of them in such a way they are representatives of the SHGs and women not of the SHGs (Plates 4.1 to 4.3). The sample does have a discernible stratification in regard to socio-economic of the people of the villages. In all, 300 sample respondents have been interviewed using the questionnaires. The questionnaires have however been designed for the study using several of the standardized questionnaires available from researchers across the world and has deliberately been adapted in a way that they could be relevant to the people and culture of the township and Kanchipuram district (Plate 4.4). Each questionnaire has taken anywhere between an hour and 90 minutes to fill-in by the women while administered orally and free associationally.

Plate 4.1: Women Self-Help Groups across Tamil Nadu, in animated discussions
Plate 4.2: A group of womenSHG members listening intently to the scholar about her research and discussing about household economic development.

Plate 4.3: Self-Help Groups in Action and Livelihoods in Nandivaram-Guduvancheri, Chennai
The Questionnaire: Construction of Research Tool

As per the research design, and for the purpose of data collection, the investigator has constructed two interview schedules. The custom-prepared questionnaires have been subjected to jury opinions. Based on the jury opinions, some items of the first drafts of the questionnaires have been deleted and some others have been modified and finally the interview questionnaires have been streamlined for use. Nearly a man-month of work has been expended on the questionnaires to make them relevant to the study and to the women of the study area.

Plate 4.4: Periurban, Chennai, Sky Line Change
Gaining Access

The one problem has indeed been in gaining access to women respondents of the town, and the problems of being able to study them, and gain some familiarity with their world, or share their ‘realities’. Social science methodologies propose many different ways in which this can be achieved, and guidance on the extent to which involvement in a respondent’s *life world* is necessary to go for particular descriptive or analytical ends. Practical problems include getting people to speak to the researcher at all, and then getting them to be open, co-operative, and sincere in discussing the aspects of their lives that interest the researcher. More fundamental for the research methodology has been the reliance on the personal descriptions of the respondents, who may either intentionally conceal or mislead the researcher, or unintentionally mislead them. In studying a woman’s *life world*, it is unlikely that the respondent will be able to comprehensively and thoroughly describe not only her opinions and thoughts, but the details of everyday activities and relationships, and the context in which they conduct them, especially in the space of a relatively short interview such as the one the researcher has been involved with each one of the respondents in Nandivaram-Guduvancheri town.

Participant observation is a method that tries to surmount these obstacles, but at the expense of huge effort by the researcher, and can only be carried out in a situation where the researcher can actually live or work within a small group over an extended period of time. Since the researcher has intended to look at a number of groups, and across the social network of one particular member, this type of methodology has been rather difficult, if not impossible. What is more, gaining access to the work place or social clubs may have been possible, but living in private homes is very difficult. Only a few researchers have tried this: for example, James Lull in order to examine media use (Lull, 1990).

Pilot Study

A pilot study was conducted using the two questionnaires, each with 30 women in a town precinct nearby and not in Nandivaram-Guduvancheri town. This was done essentially to test the relevance and adequacy of the questionnaires for collecting data towards assessing household economic development in households
of women who are active members of self-help groups operating in the town and in households of women who are not members of any self-help group, now or before. Proper instructions were given to the women of the two groups before the administration of the questionnaires, separately and individually. This has enabled the scholar to identify the vague, or ambiguous or difficult to understand questions and replace them with appropriately worded, easy-to-understand questions in the final questionnaires used with the women samples of Nandivaram-Guduvancheri later.

The Questionnaires

Given below are descriptions of the details of data solicited from the women samples using the two questionnaires, by sections of the questionnaires.

The Questionnaire for Women of the SHGs

This questionnaire was designed with 8 different sections, each focusing on a particular aspect of women interviewed, SHGs and household economic development as perceived by the women of the SHGs. Personal details of the women respondents were from the first 10 questions, in four different sections, namely: (a) personal information, (b) household / family details, (c) occupations and sources of income, and (d) expenditures on consumption and related aspects. The other four sections deal particularly with (e) details of assets, moveable and immoveable, (f) details about SHGs, (g) details on credits, especially micro-credits, and (h) development. While the sections (a) to (g) are simple questions soliciting details and information, often in quantities, the last section on development is a set of statements which are scaled statements, using a 10-point Likert type of scaling: high (5), medium (3), and low (2). Aspects for scaling are: social change (10 items), economic change (7 items), change in health (5 items), awareness of law (5 items), and empowerment (6 items). There is also a general option which solicits information as to whether the women households have become better off or worse off, economically.
Personal details collected relate to the SHG the respondent belongs to, her age, marital age, her current occupation, other jobs held, and educational attainment including the highest educational attainment. Household / family details include family size, number of adult men, adult women, boys and girls in the family, and the economically active men and boys, and women and girls. Occupational and sources of income details include just two information, one, primary occupation of the respondent (agriculture, pastoral activities, industrial work, business and services industry) and the other income sources (wages, rent, interest from deposits, petty businesses, and salary).

Details on expenditures solicited relate 9 different items of expenditures on consumption, socio-cultural events, loan repayments, travel expenses, investments made, and savings of different kinds (bank, chit, LIC, and SHG). As for details of assets, moveable assets such as small and large livestock, poultry birds, gold jewels, silver, vehicles such as two wheelers and three wheelers, and home appliances such as television, refrigerator, mixie and grinder and immoveable assets such as housing and lands are included. Depending on the nature and size of assets owned, household status in economic development may be measured just as expenditures on different items could give us an idea of resources at the household’s disposal and also the conservation and preservation of resources by the households for the future.

Details on SHGs relate to sources of assistance and loans to the SHG (NABARD, SGSY, RMK and others), reasons for establishing the SHG (high family expenditure, poverty, unemployment, emergency needs, availing loans, easy mode), functions of the SHG (savings, helping people with credits, capacity building among women, and marketing products) and the training given by the SHG (such as embroidery, tailoring, cottage and craft, art and skill and other strategies) (Plate 4.5). The details collected in regard to credits are: sources of credits (bank, SHG, and other lending institutions), money borrowed, time taken to sanction the loan, purpose of loan (for start-ups, price rise, children’s education, medical treatment), repayment of loan, and placement of trainees on jobs.
The questionnaire also has 38 items for scaling under variable heads of development or change. Development and/or change in the households is for the current period and is rated against the pre-SHG period. Under social change, 10 items of scaling are included and they are, in a nutshell, reading and writing with good skills, knowledge of bank transactions, life better than before, increase in respect for women in the households, family members’ acceptance of decisions by women of the household, increased independence for women, increase in women’s self-confidence, increase in women’s outside movements, development of their personalities, and the development of leadership in women. There are 7 items for scaling under economic change and they are the ability of women to contribute to family income, ability to save through SHG, ability to invest money in savings through banks, reduced poverty and hardships, improved economic status now against the pre-SHG period, decrease in dependence on spouse for family expenses, and ability to bear the cost of children’s education (Plate 4.6). Improvements in household / family economics are expected to bring improvements in health care and so the items scaled under health change are: clear / clean water facility, reduced malnutrition, awareness about health problems, consciousness about family planning, and the visits to PHCs for treating illnesses.
Plate 4.6: Two examples of Livelihoods SHG women are engaged in, making a living

There are 5 items under environmental change and they relate to women’s knowledge about water management, their knowledge about proper disposal of garbage, their awareness of environmental cleanliness and sanitation, electrified streets, and knowledge about pollution. Awareness of legal rights of women is an area under awareness of law and there are 5 items for scaling here as well: women’s awareness about women’s rights, about equality of men and women, knowledge of divorce and remarriage, knowledge about property inheritance, and knowledge about prohibition of dowry. Finally, there are 6 items for scaling under women empowerment and they are awareness of science and technology, developed skills of problem-solving in women, active role for women in the family, women’s leadership qualities better than before, proper knowledge about voting and electing a correct person, and women’s ability or willingness to contest local elections.
Additionally, there is a question as to whether the households of women have become better off than before (their becoming members of SHGs. Overall, this questionnaire has the ability to generate as many as 90 pieces of information, most of which are quantities and some of which are qualitative.

**The questionnaire for women of the town who are not members of the SHGs:**

This questionnaire was designed with 8 different sections, each focusing on a particular aspect of women interviewed and household economic development. Personal details of the respondents solicited were from the first 13 questions in four different sections, namely, (a) personal information, (b) household / family details, (c) occupational characteristics, and (d) income from different sources. The next four aspects are on different and specific details: (e) expenditure on consumption items and socio-cultural and financial details, (f) details of assets, both immovable and moveable, importantly in monetary terms, (g) problems and constraints in socio-economic development faced by the people as seen through the eyes of women of the SHGs, and (h) difficulties in production and exchange (marketing) and related aspects of economics.

This questionnaire is designed in a way it is somewhat different in content than the questionnaire for women of the SHGs. Comparisons to a certain extent are possible between the two databases, insofar as we may assess household economic development in households of women of SHGs and those of women who are not members of SHGs. In this questionnaire, personal information is gathered from the first 7 questions: name, age, educational attainment, highest level of education attained, occupation of the spouse, respondent’s occupation, and secondary occupation if any. Household / family details and occupational and income details solicited here are more or less similar to the ones in the questionnaire for women of the SHGs. But the details on occupation such as the length of period in the present occupation, hours of work, description of work, willingness to switch to another job and the reasons for the same, opportunities and skills available for women are some of the pieces of information solicited here are additional, more incisive. Income from different sources – main, auxillary occupations – its adequacy or otherwise are again detailed information from the questionnaire. Even expenditures on
consumption, details on assets give us greater details than the questionnaire for women of SHGs. Problems and constraints in socio-economic development relate to transport access, distance to workplaces, costs of travel by different modes, and women are asked to rate their technical and professional knowledge and also their skills using a 10-point Likert scaling.

Difficulties in production and exchange are a section soliciting information, in fact scaled ratings, on raw materials availability, land-space-accommodation, capital, and market. There are pieces of scaled data on women’s livelihoods and hardships in several facets of development: poverty, mal- and under-nutrition, low quality of life, low hygiene and sanitation levels, lack of help from friends and relatives, discrimination by spouses, dependence, and crisis situation in children’s education. The questionnaire further seeks solutions from the women as they face several problems in socio-economics of their living even as they are asked to point out to who is responsible for the difficulties in their livelihoods. Overall, the questionnaire for women who are not the members of the SHGs is capable of giving us a database with as many as 85 or so good and different variables.

The two databases could be used to first understand the status of women, in SHGs and outside of them, and then their plight in making a living and the juggle they have to perform in their day-to-day life and work towards the betterment of their households, and also towards building a future for their children in the ever prevalent difficult circumstances in the town of Nandivaram-Guduvancheri.

Reliability and Validity of the Dataset

The reliability of the interview questionnaire has been established by calculating the Cronbach Alpha ($r = 0.75$) and the intrinsic validity was established by taking the square root of the reliability co-efficient $C = 0.75$; that is, $r = 0.84$. Thus, from the two coefficients, it may be inferred that this tool is highly reliable and valid.

Secondary Data Sources

The secondary sources of data have generally been census abstracts of various census years, documentary sources of the government agencies, research
reports and international and national reports on SHGs and women by different organizations such as the UN, International Population Institutes and national institutions of different kinds. Data as well as perspectives have been collated from the reports and documents for the purpose of writing up the thesis.

**Statistical Methods of Analysis**

Among the statistical tools used in the study are (a) the simple frequency and percentage analysis of questionnaire survey data (one-way as well as two-way tables) and (b) the multivariate statistical analysis of factor using principal components approach. In order that the data are amenable to statistical analyses, the questionnaire data have been converted into 2 datasets using the MS EXCEL spreadsheet and the analyses themselves have been performed using the SPSS package. The two methods are described in some detail below:

**Simple Frequency and Percentage Analysis**

For the purpose of description of sample and respondent related characteristics, a frequency and percentage analysis has been done for all variables extracted from the questionnaire and put into the dataset. First, a simple frequency of each of the fields with column percentages has been made and then two-way tables using certain select pairs of variables have been carried out, in order to measure variations. The analyst begins to explore the data, by measuring the central tendency of the data, and more importantly, the dispersion of the data around this central tendency.

Frequency analysis is particularly useful for describing discrete categories of data having multiple choice or yes/no response formats. This analysis involves constructing a frequency distribution. The only technical requirement of the frequency analysis is that the categories of response be mutually exclusive and exhaustive. This means that the same observation cannot be counted as belonging to more than one response category. The frequency analysis must be exhaustive in the sense that all respondents must fit into a category. The tables so generated are numerous, only select tables are therefore included in the text while others are interpreted so as to show the variations therein.
Common Factor Analysis

Factor analysis is a statistical technique designed to analyze the interrelationships within a set of variables by reducing the complex data to an easily interpretable form (Davis, 2002). In multivariate analysis, the bi-variate techniques are extended so that more than two variables can be considered, the ‘m’ variable becoming the ‘m’ axes of the test space. Procedures of multivariate analysis are often concerned with the problem of reducing the original test space to the minimum number of dimensions needed to describe the relevant information contained in the original observations. Multivariate procedures differ in the types of original information they preserve. Some understanding of matrix algebra is essential to using and understanding the multivariate analysis.

It is a particular psychometric model that has been in wide use in social sciences. This helps in the study of the logical implications of systematic intercorrelations within sets of tests. However, the social sciences follow just one of the many approaches to the reduction of dimensionality in correlated systems of measurements and the rotation (varimax, a short form for maximizing variance, for example) of a reduced number of axes to more meaningful positions.

The Factor Analysis (FA) is also a classification procedure in that it may be usefully applied to multivariate situations to classifying the N individuals, on the basis of ‘m’ variables. One particular feature of the FA is that ‘p’ underlying factors in the multivariate sample space model is always less than the ‘m’ variables: that is, \( p < m \). The underlying factor dimensions are drawn from the use of intercorrelations system by generating ‘p’ number of scores each for the ‘N’ individuals. The scores may however be drawn from the varimax rotation, which stands for maximizing variance.

The purpose of factor analysis is to interpret the structure within the variance-covariance matrices of the multivariate data collection made on the different aspects of household economic development in Nandivaram-Guduvancheri town panchayat, in Tamil Nadu, India. The basic mathematical operations in factor analysis may be stated as follows:
\[ Z_j = a_{j1} P + a_{j2} P_2 + \ldots + a_{jm} P_m \] where
\[ Z_j = X_j \cdot X_{\text{mean}} / O_j \] or standardized variable
\[ P_i = (i = 1, 2, \ldots, m) \] are the principal components and
\[ a_j = (j = 1, 2, \ldots, n) \] are the coefficients or factor loadings of \((i = 1, 2, \ldots, m)\) \(j^{\text{th}}\) variable relating to the \(i^{\text{th}}\) component.

In other words, each factor is nothing but a linear combination of weighted variables which can also be expressed as:

\[ P_1 = a_j X_j \] where

Thus, in factor analysis, a data matrix containing measurements on ‘m’ variables for each of ‘N’ observations is analyzed.

The technique uses extraction of the eigen values and eigenvectors from the matrices of correlations or covariance. The basic mathematical operations in factor analysis are done with many embellishments on the procedures.

FA is a deep and complex methodology. It is one of the most widely used multivariate procedures. The model is based on several unique assumptions. For one, \textit{the precise number of factor is assumed prior to the analysis}. The factors extracted, or rather the number of factors, are validated by the variance each of them explain to the total. There is a progressive decline in the value of variances with the increasing number of factor dimensions. The first or the main factor dimension has the highest of the total variance explained and the bipolar the next highest and so on, resulting in progressively declining variance.

The FA requires that ‘p’, the number of factors, be known prior to analysis. This implies that the \textit{investigator has some insight into the probable nature of the factors} and can predict a suitable number of factors to be extracted (see Figures 4.1 and 4.2).
The eigen value operation in factor analysis is performed on a standardized variance-covariance or correlation matrix. Hence, the FA used here is said to be R-mode factor analysis. This assumes not only that all variables are weighted equally, but also allows us to convert the principal component vectors into factors. In larger matrices such as ours, the eigenvalues usually are more uniform for standardized data than for raw data. And to perform the FA, it is necessary that we convert our unit, or normalize eigenvalue. The result is a factor, a vector, which is weighted proportionally to the amount of total variance it represents.

Figure 4.1: Factor analysis: Analyze and select in SPSS
The elements in the factors are referred to as factor loadings. The eigenvalues represent the proportion of the total variance accounted for by the eigenvectors. The factor loadings on the other hand are the correlation values between the old and the new, transformed variables.

If we arrange the factor loadings in a matrix form, we have then a factor matrix. If we square the elements in the factor matrix and sum within each variable, the totals are the amount of variance of each variable retained in the factors. These sums are referred to as the communalities and are symbolically represented as $h_j^2$. The communalities are equal to the original variances.

A specific rule that most factor analysts suggest in the extraction of factor is that of retaining all factors, which have eigen values greater than one. That is, retain all factors, which contain greater variance than the original standardized variables. But of course in most instances only a few of the factors will contain most of the variances in the dataset and hence this recommendation is useful. If factor theory is applicable to any given dataset, a few factors should account for a very high
percentage of the variance and the communalities of the variables found under each factor dimension is high.

The FA is said to be *reducing the dimensionality* of a problem to a manageable size. However, the meaning of the factors may be difficult to deduce. This problem is overcome by resorting to maximization of the variance of the loadings on the factors. This in other words is maximizing the range of the loadings. This is done in the analysis here by a rotation procedure called Kaiser’s varimax rotation. The rotation of the factor axes is performed, iteratively. The analysis also results in factor scores, which represent estimates of the contribution of various factors to each original observation (women of the SHGs and those who are not members of the SHGs). In fact, factors themselves are estimated from these same data. Thus the computation of factor scores is somewhat a circular process and the results are not unique. Factor analysis explains in a sense the interrelationships in a large number of variables by the presence of a few factors (Kaiser, 1958; Harman, 1960; Lawrence and Upchurch, 1983).

The factor extraction is done with a minimum acceptable eigen value of >1.0 (Kaiser, 1958; Harman, 1960). The factor loading matrix is rotated to an orthogonal simple structure, according to varimax rotation. It results in maximization of variance of factor loadings of the variables. This procedure renders a new rotated factor matrix in which each factor is described in terms of only those variables and affords greater ease for interpretation. Factor loading is a measure of the degree of closeness between the variables and the factor. The largest loading, either positive or negative, suggests the meaning of the dimension; positive loading indicates that the contribution of the variables increases with the increasing loadings in a dimension; and negative loading indicates a decrease (Lawrence and Upchurch, 1983).

Factor analysis manages over a hundred variables, compensates for random error and invalidity, and disentangles complex interrelationships into their major and distinct regularities. It is mathematically complicated and has diverse and numerous considerations in application. The purpose of this discussion is to enhance the understanding and utilization of the results of factor analysis, rather than provide a technical description.
Conceptually, factor analysis is a means by which the *regularity* and *order* in phenomena can be discerned. As phenomena co-occur in space and in time, they are patterned; and as these co-occurring phenomena are however independent of each other, there are a number of *distinct patterns* (Velicer and Jackson, 1990). What factor analysis does is this: it takes measurements and qualitative observations and resolves them into distinct patterns of occurrence. It makes explicit and more precise the building of fact-linkages going on continuously in the human mind.

Factor analysis applied to delineate patterns of variation in characteristics is called R-factor analysis. It is applied in order to explore a content area, structure a domain, map unknown concepts, classify or reduce data, illuminate causal nexuses, screen or transform data, define relationships, test hypotheses, formulate theories, control variables, or make inferences (Rummel, 1970; Ram, 1982; Davis, 2002).

When a table of data, say, answers to a questionnaire, are interrelated in a complex fashion, then factor analysis is used to untangle the *linear relationships* into their separate patterns. Each pattern appears then as a *factor* delineating a distinct cluster of interrelated data. It is useful for reducing a mass of information to an economical description. For example, the data are facilitated by reducing them to their *common factor patterns*. The factors concentrate and index the dispersed information in the original data and can therefore replace the characteristics without much loss of information.

Factor analysis is often employed to discover the basic *structure* of a domain. It can be used to group interdependent variables into descriptive categories. It is also used to classify, for example, individual profiles into types with similar characteristics or behaviour. Or it is used on data matrices of a social-choice type to show how individuals, or social groups cluster on their transactions with or choices of each other.

An investigator often wishes to develop a *scale* on which individuals, or groups, or even a community can be rated and compared. One problem in developing a scale is however to weight the characteristics being combined. The analysis thus offers a solution by dividing the characteristics into independent sources of variation (factors). Each factor then represents a scale based on the
empirical relationships among the characteristics. The factor score results are actually such scales, developed by summing characteristics times the weights (see Distephano, Zhu and Mindrila, 2009).

Hypotheses abound regarding dimensions of attitude, personality, group, social behaviour, and revealed perceptions. Since the meaning usually associated with ‘dimension’ is that of a cluster or group of highly inter-correlated characteristics or behaviour, factor analysis is used to test for their empirical existence. Which characteristics or behaviour is, by theory, related to which dimensions can be postulated in advance and statistical tests of significance can be applied to the factor analysis results (Steiger, 1990).

Common factor analysis is concerned with defining the patterns of common variation among a set of variables. Variation unique to a variable is however ignored. In contrast, another factor model called component factor analysis is concerned with patterning all the variation in a set of variables, whether common or unique (see Kaiser, 1958; Harman, 1960; Lawrence and Upchurch, 1983; Vinayakam and Sekar, 2013).

The Algebraic Model

A traditional approach to expressing relationships is to establish the mathematical function f(X, W, Z) connecting one variable, Y, with the set of variables X, W, and Z. Such a function might be \( Y = 2X + 3Z - 2W \), or \( Y = 4XW/Z \). The variables on both the right and the left side of the equation are known, data are available, and it is only a question of determining the best function for describing the relationships. Let us assume that there are several Y variables and our Y variables are related to a number of functions operating linearly. That is,

Equation 1:
\[
Y_1 = z_{11}F_1 + z_{12}F_2 + \ldots + z_{1m}F_m, \\
Y_2 = z_{21}F_1 + z_{22}F_2 + \ldots + z_{2m}F_m, \\
Y_3 = z_{31}F_1 + z_{32}F_2 + \ldots + z_{3m}F_m, \\
\ldots \\
\ldots \\
\ldots \\
Y_n = z_{n1}F_1 + z_{n2}F_2 + \ldots + z_{nm}F_m,
\]
where:

\[ Y = \text{a variable with known data} \]

\[ a = \text{a constant} \]

\[ F = \text{a function, } f(\text{ of some unknown variables).} \]

This is important in understanding factor analysis to remember that \( F \) stands for a function of variables and not a variable. For example, the functions might be \( F_1 = XW + 2Z \), and \( F_2 = 3X^2Z/W^{1/2} \). The unknown variables entering into each function, \( F \), of Equation 1 are related in unknown ways, although the equations relating the functions themselves are linear.

Within the algebraic perspective, what does factor analysis do? By application to the known data on the \( Y \) variables, \textit{factor analysis defines the unknown \( F \) functions}. The loadings emerging from a factor analysis are the \( a \) constants. The factors are the \( F \) functions. The size of each loading for each factor measures how much that specific function is related to \( Y \). For any of the \( Y \) variables of Equation 1 we may write:

\text{Equation 2:}

\[ Y = a_1F_1 + a_2F_2 + a_3F_3 + \ldots + a_mF_m, \]

with the \( F \)'s representing factors and the \( a \)'s representing loadings.

It is possible that some of the \( F \) functions are common to several variables. These are called \textit{group factors} and their delineation is often the goal of factor analysis. Besides determining the loadings, \( a \), factor analysis also generates data (scores) for each case (individual, migrant in our case) on each of the \( F \) functions uncovered. The derived values for each case are called \textit{factor scores}. They, along with the data on \( Y \) and Equation 1 give a mathematical relationship among data as useful and important as the classical equations like \( Y = 2X + 3Z \).
Graphical and Other Representations (SPSS, MSEXCEL)

Graphical representations and charts have been created to illustrate the questionnaire survey data analyzed. Maps used in the thesis have generally been prepared using the modern geographical information systems software.

Library Research

In addition library research has been gone through meticulously using different University libraries, especially University of Madras, and also other research institutions such as the Madras Institute of Development Studies of the University of Madras and the Central library of the district under study.

Conclusion

This chapter has dealt with the research methodology adopted in the study of Self-Help Groups and Household Economic Development with a case of SHGs in Nandivaram-Guduvancheri town panchayat, particularly for a sample of 300 women chosen from select Self-Help Groups in the town and also women from the town who are not members of any SHGs. The discussion in the chapter has focused on the qualitative and quantitative research methodologies that could be useful in any study and then narrowing the focus on the five areas of a methodology for the present study, namely, the primary sources of data, the secondary sources of data, the statistical techniques used in the study using simple frequency and percentage analysis and multivariate common factor analysis, graphical representations using the SPSS and MS EXCEL software and also the mapping software GIS, and library research involved in the study. The methods chosen for the current research are most appropriate to assemble ideas, illustrate data and analyse, interpret and infer conclusions from the analysis of questionnaire survey data. In the next chapter, the discussion turns essentially to the household economic development dimensions of women’s self-help groups of Nandivaram-Guduvancheri and also focuses on the dimensions of household economic development in the town of our study.