Chapter III
METHODOLOGY

The methodology adopted in the current study is “Health and Sanitation Scenario in Selected Urban Slums –A Micro Level Study” discussed under the following heads:-

3.1 Locale of the Study

3.2 Selection of Sample

3.3 Data base of the Study

3.4 Period of Study

3.5 Techniques of Analysis

3.6 Limitations of the Study

3.1 Locale of the study

Coimbatore, popularly known as the Manchester of South, is the third largest city and the second among the most industrialised cities in the world. It is located in the western part of Tamil Nadu, on the banks of river Noyyal. It is surrounded by the Fairy Queen; The Nilgiris (the Blue Hills) in the north, the revolutionary Western Ghats side of Kerala in the west, newly formed Tiruppur in the south and south east, and the highly agriculturally commercial turmeric Erode District in the East. This highly progressive, entrepreneurial and commercial district of Tamil Nadu lies between 10, " - 10' and 11," - 30' Northern latitude and 76,"-40’ and 77,"-30’ Eastern longitude. The district has a geographical area of 7469 sq.kms. The district is divided into three revenue divisions, 9 taluks, 19 blocks and 482 revenue villages.

According to 2011 Census, Coimbatore had a population of 1,050,721 with a sex-ratio of 997 females for every 1,000 males, much above the national average of 929. A total of 102,069 were under the age of six, constituting 52,275 males and 49,794 females. The average literacy of the city was

82.43 percent, compared to the national average of 72.99 percent. There were a total of 425,115 workers, comprising 1,539 cultivators, 2,908 main agricultural laborers, 11,789 in household industries, 385,802 other workers, 23,077 marginal workers, 531 marginal cultivators, 500 marginal agricultural labourers, 1,169 marginal workers in household industries and 20,877 other marginal workers. Located in the rain shadow region of Western Ghats, Coimbatore enjoys pleasant weather throughout the year. The rich red loam soil and red sandy soil in the district are favourable for production of cotton and a wide variety of cereals and food grains, spices, and condiments. The region has a total cultivable area of 330,584 hectares. Forest coverage spans across 158,801 hectares and is primarily suitable for timber, mango, walnut, and silk cotton.

Selection of sample

Coimbatore Corporation is divided into five administrative zones East, West, North, South and Central zone and each zone is further divided into e 20 wards. Details of the slums in Coimbatore (both notified and non-notified) have been collected from the Coimbatore Corporation through Tamil Nadu Slum Clearance Board. The slum population of constitutes nearly 16 percent of the total population of the Corporation. Emergence of IT, Educational institution, Industrial growth, and augmentation of medium and small scale industries aggravate many to the city. High rents, prohibitive land prices, very limited or no access to credit facilities for the urban poor, and lack of credit worthiness have been the important factors contributing to the growth of the slums in Coimbatore Corporation area. Estimated statistics from the socio economic survey of Slum Clearance Board shows that there are 29937 slum households in Coimbatore Corporation of 3.42 sq.km area. The investigator selected the slum which had the highest population from each zone. Teenability analysis done by the slum clearance board indicates the slums that falls in the status of fitness of land for human habitation. Those households which lack private toilets were drawn and from among them 30 percent of the households were selected. The data pertaining to the study was exclusively collected from the women folks only. The sample size from each zone is given below:

TABLE 3.1

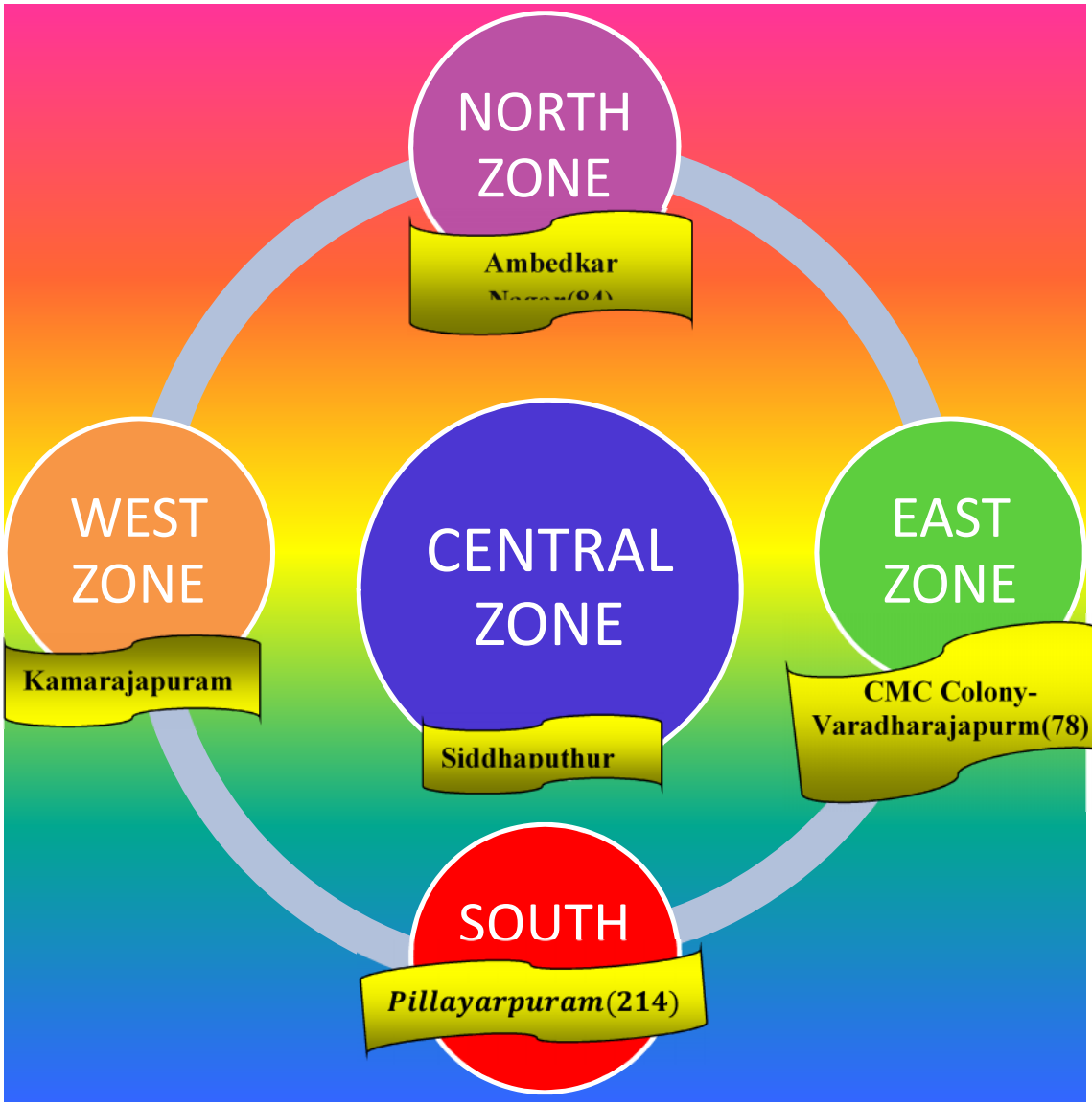
SELECTED SAMPLE

Zone	Ward No	Ward	No of Slum Households	Sample size proposed for survey	Sample size actually surveyed
North Zone	026	Ambedkar Nagar-Meenatchi Amman Koil Veethi	279	84	76
East Zone	064	CMC Colony-Varadharajapuram	260	78	74
West Zone	024	Kamarajapuram	639	192	180
South Zone	097	Pillayarpuram	714	214	200
Central Zone	052	Siddhaputhur part 1&2	667	200	193
Total				768	723

From these selected areas 768 samples were selected by adopting incidental purposive sampling technique. The term incidental sampling is applied to those samples that are taken because they are most readily available. The basic assumption behind purposive sampling is that with good judgment and an appropriate strategy one can handpick the cases to be included in the sample and thus develop samples that are satisfactory in relation to one's needs (Guilford, 1978). A common strategy of purposive sampling is to pick up cases that are judged to be typical of the population, in which one is interested, assuming that errors of judgment in selection will tend to counter balance each other if sufficiently large sample is taken. Questionnaire was administered by the investigator within the selected area of the city. Every household was asked to complete the questionnaire after identifying whether they were in the slums as mentioned by slum clearance board. A total of 768 women were contacted. The complete information were collected from 723 women only, as 45 of them were not interested in giving data.

FIGURE 3.1

SELETION OF THE SAMPLE



3.3 Data base of the study

Data pertaining to the study were collected by personal interview method. The interview schedule consisted of questions relating to the demographic profile of the slum dwellers, their living conditions and family background, working conditions, health conditions, the problems faced etc. The schedule was pre-tested and based on their responses the questions were reformulated and the final interview schedule was framed.(Annexure I).

3.4 Period of the Study

Data for the study were collected from the sample units by administering a pre-tested interview schedule during the period May 2015-December 2016.

3.5 Techniques of Analysis

Besides averages, percentages and graphs, the following techniques were applied.

3.5. 1 The Multinomial logistic analysis

Modelling of risk processes such as risk awareness, risk identification, monitoring and reporting, planning and mitigation etc is among rather difficult subjects tackled by risk analyst especially in applying multinomial logistic regression in dynamic (social) setting. Invariably though, social science research (Yu, Lai and Wang, 2008) problems somewhat call for analysis and prediction of a dichotomous1 outcomes. Traditionally, such research outcomes were addressed by either ordinary least squares (OLS) or linear discriminant function analysis (Hosmer and Lemeshow, 2000). However, both techniques, as a result of their nature, depend on strict statistical assumptions, thus, normality of independent variables, linearity of relationships, multi collinearity among independent variables, equal dispersion matrices for discriminant analysis (Tabachnick *et al*., 2001). These assumptions which are not easily observed in a dynamic setting are part of multiple regression. Introduction of multinomial logistic regression was an alternative regression analysis to cater for conditions that do not necessarily obey the assumptions listed above with the exception of multicollinearity (Hosmer and Lemeshow, 2000).

The logistic regression model is shown as the following form:

$$\ln(p/1-p) = \beta o+\beta iXi$$

where,

p=the probability of resource substitution behaviour;

(p/1-p) = odds of resource substitution behaviour;

βo = constant;

Xi = vector of independent variables;

βi = parameter estimate for the ith independent variable.

The logistic regression is powerful in its ability to estimate the individual effects of continuous or categorical independent variables on categorical dependent variables (Wright 1995).

The multinomial logistic regression model used is generally effective where the dependent variable is composed of a polytomous category having multiple choices. The basic concept was generalized from binary logistic regression (Aldrich & Nelson 1984, Hosmer & Lemeshow 2000). In a multinomial logistic regression model, the estimates for the parameter can be identified compared to a baseline category (Long, 1997). In this study, having no willingness to substitute was specifi ed as the baseline category. The multinomial logistic regression model with a baseline category would be expressed as follows:

$$\text{Log } (\pi i/ \pi l) = \alpha i + \beta ix,\ i = 1,\ldots..,I-1.$$

The logistic model uses the baseline-category logits with a predictor x. This multinomial logistic regression model can be a useful tool for modeling where the dependent variable is a discrete set of more than two choices.

3.5.2 Garret Ranking scale

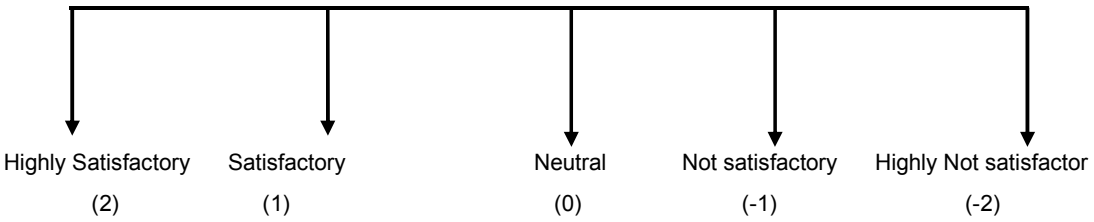
Garrett's rating scale is applied to find out the strength of factors ranked by the respondents in relation to the hygiene practices followed. From the ranks given for each factor, percent positions were calculated by using the formula.

Percent position = 100 *(R-0.5)/N

where R is the rank assigned and N is the number of items ranked. The percent position was then converted into scores using Garrett's scores table (Garrett H, 2005). Garret ranking scale technique was used in ranking the various reasons for selection of healthcare services in their order of priority.

3.5.3 Likerts' Summated Scale

The Likert summated scaling technique was used to scale the problems of the slum dwellers.In the Likert scale, the respondent were asked to respond to each of the statements in terms of five degrees of agreement or disagreement.Each point on the scale carries a score. Response indicating the least favourable degree of satisfaction is given the least score (1) and the most Highly satisfactory (2),Neutral (0) Not satisfactory (-1),Highly not satisfactory (-2).



These score values are normally not printed on the instrument but are shown here just to indicate the scoring pattern. The Likert scaling technique, thus, assigns a scale value to each of the five responses. The same procedure is repeated for each and every statement in the instrument. This way the instrument yields a total score for each respondent, which would then measure the respondent's favourableness toward the given point of view.

3.5.4 Cronbach's Alpha

Cronbach's alpha evaluates the unidimensionality of a set of scale items. It is a measure of the extent to which all the variables in a scale are positively related to each other. In fact, it is really just an adjustment to the average correlation between every variable and every other. The formula for alpha is

$$\alpha standardized = \frac{K.\bar{r}}{(1 + (K - 1).\bar{r})}$$

where k is the number of variables and *r* is the average correlation among all pairs of variables. Cronbach's alpha values ranges from 0 to 1. The higher the score, the more reliable the generated scale is, Nunnally (1978) has indicated 0.7 to be an acceptable reliability coefficient but lower thresholds are sometimes used in the literature. In the study, the reliability testing was done for the statements pertaining to problems of women in slums.

3.5.5 Factor Analysis

Factor analysis is a generic name given to a class of multivariate technique whose primary purpose is to define the underlying structure in a data matrix. Broadly speaking, it addresses the problem of analyzing the structure of the interrelationships (correlations) among a large number of variables by defining a set of common underlying dimensions, known as factors. With factor analysis, the researcher can first identify the separate dimensions of the structure and then determine the extent to which each variable is explained by each dimension. Once these dimensions and the explanation of each variable are determined, the two primary uses for factor analysis, namely summarization and data reduction can be achieved. In summarizing the data, factor analysis derives underlying dimensions that, when interpreted and understood, describe the data in a much smaller number of concepts than the original individual variables. Factor analysis was used in the present study to identify the underlying pattern of relationship between the various problems pertaining to sanitation faced by the women in the study area ,and identifying the benefits and problems encountered in accessing better sanitation facilities.

3.5.6 Environmental Sanitation Index:

The data collected using the scheduled Questionnaire was used to formulate Environmental Sanitation Index(ESI).The indicators such as, access to drinking water, usage of toilet, condition of drainage ,solid waste collection were used to access and formulate the Sanitation Index. United Nations Development

Programme (UNDP), Balamurugan & Ravichandran (2014), Mukherjee and Chakraborty (2007) were referred for selecting the variables to be included for calculating Sanitation Index.

The ESI is composed of indicators selected from the fields of environmental sanitation, socio-economics, and public health and hydro sources. The variables chosen were the ones deemed easy to access and easy to rate, in order to facilitate the elaboration of the study. There are four groups of indicators: The variables are grouped into four categories, namely access to Drinking Water, Usage of Toilet, Condition of the Drainage , Solid Waste management .

The indicators under the variable of 'Drinking water' are availability of Piped water availability of ground water, availability of surface water adapting purification process of water .Purification of drinking water is given high score followed by piped water ,ground water and surface water. Similarly the indicators for the variable ,Usage of Toilet are ,Open defecation ,common toilets, pit latrines and septic tank latrines .Highest score is given to the separate toilet with septic tank ,followed by pit latrines ,Common toilets and Open defecation. Condition of drainage is classified into four categories , underground drainage, open pucca, closed pucca, and absence of drainage .Underground drainage is given high score followed by closed pucca and open pucca. No drainage condition is given least score of Zero.

The responses are presented with certain indicators and are measured in terms of their relative position of the composite index. The total number of indicators is 4. The minimum one can score on a particular indicator is zero and the maximum is 4. The scores were transformed into index by using the following formula:

Thus,

$$\frac{\text{Actual values}_{ij} - \text{Minimum Value}_{jit}}{\text{Maximum value}_{j} - \text{Minimum Value}_{j}}$$

Once the indicator indices are formed, the comprehensive Composite Index is then calculated as a simple average of the indicator indices.

$$\text{Composite Index} = \frac{\sum \text{Individual Indicator indices}}{N}$$

Where, composite index is the summation of all the individual indices and N is the total number of individual indicator indices.

Sanitation Index (SI) of the households were constructed by using 4 indicators namely Disposal of waste, Drainage, Drinking water and Toilet.

The classification on the basis of the total score used for analysis is as follows:

- **1-1.75 = Poor sanitation index**
- **1.75-2 .50 = Average sanitation index**
- **2.50-3.25 = Fair sanitation index**
- **3.25-4.00 = Good sanitation index**

3.5.7 Estimated Logit Model

Estimated Logit Model is also known as the logistic regression. The logistic regression is one that specifies a functional relationship between a basically dichotomous dependent variable and categorical or metric scaled variables. In fact it is a method of multivariate analysis of the multiple regression model designed to deal with the situation when one has the measurement of presence or absence, occurrence or non-occurrence of some factors. Logistic regression is concerned with modeling the odds of dependent variable and the parameters for logistic are most easily interpreted as they are expressed as odd ratios. The basic form of logistic function is:

$$P = \frac{1}{1 + e^{-Z}}$$

When numerator and denominator of the right side of the above equation aremultiplied by ez, the logistic function can be expressed in the following manner:

$$P = \frac{\exp(Z)}{1 + \exp(Z)}$$

where z is the predictor variable and e is the base of natural logarithm, equal to 2.71828. Above equation is bivariate. If z is a linear function of a set of predictor variables then:

Z = b0 + b1 X1 + b2 X2 +……+ bk Xk

This expression is substituted in the formula for logistic function. Thus, the unction becomes

$$P = \frac{1}{1 + e^{-(b0+b1x1+b2x2+\cdots\ldots+bkxk)}}$$

In this analysis, both logistic regression coefficients and odd ratios are used. Odd ratio is the ratio of the probability of the event occurring to the probability of the event not occurring and is denoted as:

Ln (Pi / 1- Pi) = b0 + b1 X1 + b2 X2 +……+ bk Xk + e

where, Pi = probability of the event occurring;

b_0 = constant term;

X_1 to X_K = independent variables;

b_1 to b_K = unknown regression coefficients associated with the independent variables X_1 to X_K.; and

e = error term representing unobserved variables that influence dependent variable.

The quantity P/1 – P is called the odds; hence the quantity Ln (Pi /1 – P) is called the log odds or the logit of P.

The coefficients are estimated using the method of maximum likelihood. The logit model was used to identify the probability of willing to pay for better sanitation services among the sample respondents.

3.5.8 Paired 'T' Test

A paired t-test is used to compare two population means having two samples in which observations in one sample can be paired with observations in the other sample. The paired t-test is used in this study to examine the impact of

educational intervention done to a sample of respondents. Data on the awareness on hygiene and sanitation before and after the intervention programme is collected and using paired t-test, the efficiency is assessed.

x = test score before the intervention,

y = test score after the intervention

To test the null hypothesis that the true mean difference is zero, the procedure is as follows:

1. Calculate the difference (di = yi − xi) between the two observations on each pair, making sure you distinguish between positive and negative differences.
2. Calculate the mean difference, $\bar{d}$
3. Calculate the standard deviation of the differences, sd, and use this to calculate the standard error of the mean difference, SE ($(\bar{d}) = \frac{s_d}{\sqrt{n}}$
4. Calculate the t-statistic, which is given by T = $\frac{\bar{d}}{SE(\bar{d})}$

5.Under the null hypothesis, this statistic follows a t-distribution with n − 1 degrees of freedom.

6. Use tables of the t-distribution to compare the value for T to the tn−1 distribution.

This will give the p-value for the paired t-test.

3.6 Limitations of the study

The present study is fundamentally based on primary data. It is obvious that primary data has its own limitations. To have accuracy in the data collected, cross checking was carried out. In this way, though inaccuracy in the given data was minimised, the data could not be considered as 100 percent correct. Further many of the women folks did not co-operate during data collection, but when the purpose was revealed they agreed and gave information to the questions pertaining to the survey. Therefore the present study relies only on the information gathered through surveys, observations and personal interviews, which are

subject to bias. As with most empirical studies, the sample size and spectrum of respondents is a limitation. Even though a concerted effort was made to include a range of different individual from different slum areas, the sample exhibited similarity in certain social habits, however in terms of personal hygiene and cleanliness there was vast differences.

Moreover the survey is not representative of the entire slums in Coimbatore. The sample was selected only from five selected zones in Coimbatore. Further, the findings and conclusion could only be applicable to a similar set of socio-economic situation. These limitations in no way negate the findings of the study, but it offers scope for further research in future.