Chapter III

Data Base and Methodology

This chapter outlines data base and methodology commensurate with the objectives of the study keeping in view the observations of Downey and Ireland (1979) that methodologies are neither appropriate nor inappropriate until they are applied to a specific research problem. It is split up in two sections. Section 3.1 contains the information as regards data base used for the study whereas section 3.2 illustrates the statistical tools and techniques employed for data analysis. Last section deals with the certain limitations of the study.

3.1 Data Base

The present study relates to public life insurance company namely, Life Insurance Corporation of India (LIC). The information regarding private life insurance players has been gathered only to gauge the overall impact of changing scenario in the life insurance sector. This study has been confined to LIC divisions (see table 4.2) and branches operating in Amritsar, Jalandhar, and Ludhiana, in Punjab (a progressive state of India). Table 3.1 shows the number of cities and corresponding number of branches established in Punjab.

<table>
<thead>
<tr>
<th>Cities</th>
<th>Number of Branches</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amritsar</td>
<td>5 (i.e. Amritsar-I, II, III, IV and CAB)</td>
</tr>
<tr>
<td>Jalandhar</td>
<td>7 (i.e. Jalandhar- I, II, III, IV and CAB, industrial area branch, P&amp;GS)</td>
</tr>
<tr>
<td>Ludhiana</td>
<td>8 (i.e. Ludhiana- I, II, III, IV and CAB, DAB, P&amp;GS )</td>
</tr>
<tr>
<td>Total</td>
<td>20</td>
</tr>
</tbody>
</table>

The database of the present study comprises of both primary and secondary sources. To study the profile of LIC, emerging competitive scenario of life insurance sector in India, and the existing marketing strategies of LIC, data has been collected from different
published sources viz., LIC annual reports, IRDA annual reports, IRDA handbook, socio-economic profile and statistical year books, journals, magazines, agents’ manual, pamphlets of LIC products, agents’ digest-cum ready reckoner (for LIC premium), books, diaries of LIC (Mumbai) as well as insurers’ websites. For the emerging competitive scenario, a period of nine years i.e. from 2001-02 to 2009-10 has been taken for data collection. For the marketing strategies, data has been collected for a period of fourteen years i.e. from 1995-96 to 2008-09. However, at certain places the period is less than ten years because of non-availability of data.

### 3.1.1 Universe of the Study

Universe of the study comprised of customers and agents of LIC from the major cities of Punjab, namely Amritsar, Jalandhar and Ludhiana. These cities were chosen because of their importance on the map of Punjab. In this state, there is a large number of LIC customers and agents. In view of the large numbers involved, it was almost impossible to cover all the customers and agents of LIC for the purpose of survey. Hence, Judgment-based-Convenience sampling method was used to conduct surveys of the customers and agents of LIC. Before collecting data from respondents, primary drafts of the questionnaires were pre-tested. Based on the written and verbal comments, some items were re-worded to eliminate ambiguity and some were deleted.

#### 3.1.1.1 Sample and Sampling Design

Sample size refers to the number of elements included in the study. As a rough guideline, there should be at least four or five times as many observations (sample size) as there are variables (Malhotra, 2007, pp 368, 643).

Surveys and questionnaires are the most commonly used method for data collection in the study of organizations (Hinkin, 1995). According to Emory and Cooper (1991) surveys are the appropriate methodology when information is needed on perceptions or attitudes and on past events. For the purpose of collecting primary data from LIC customers and agents, two separate sets of well structured questionnaires were designed and administrated on the respondents. Both the questionnaires were drafted on the basis of relevant literature. A sample of 450 customers and 350 agents from the three cities
namely, Amritsar, Jalandhar and Ludhiana was taken. Out of 450 customers, 51 questionnaires were not responded, 62 questionnaires had some incomplete responses and thus the effective sample size was 337, thereby yielding a response rate of 75 percent. Respondents were approached personally at their work place and sometimes at their residence. Out of 350 samples of agents, 92 did not respond, 33 questionnaires had some incomplete responses, so 225 questionnaires were completed in all respects and found valid responses (approximately 64 per cent response rate) for the purpose of analysis. In the present study, a sample of 337 customers and 225 agents has been taken which is five times more than the variables used. The survey of agents was conducted during May 2008 to September 2009 where as the survey of customers was conducted during June 2009 to December 2010.

Table 3.2: Geographic Sample Distribution of Customers and Agents

<table>
<thead>
<tr>
<th>Cities</th>
<th>Planned Sample of Customers</th>
<th>Valid Responses</th>
<th>Planned Sample of Agents</th>
<th>Valid Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Amritsar</td>
<td>200</td>
<td>146</td>
<td>150</td>
<td>92</td>
</tr>
<tr>
<td>Jalandhar</td>
<td>150</td>
<td>114</td>
<td>100</td>
<td>64</td>
</tr>
<tr>
<td>Ludhiana</td>
<td>100</td>
<td>77</td>
<td>100</td>
<td>69</td>
</tr>
<tr>
<td>Total</td>
<td>450</td>
<td>337</td>
<td>350</td>
<td>225</td>
</tr>
</tbody>
</table>

3.1.1.2 Construction of Questionnaire for Customers

To operationalise customers’ perception towards service quality, the present study has used a set of questionnaire (Annexure I) that sought to evaluate different aspects of the service quality. The questionnaire was divided into three sections. In section A, questionnaire consists of two scales, titled perceptions and expectations. The perceptions scale is a set of 52-items encompassing the five dimensions that describe what the customers actually think of the service. The expectations scale is a set of 52 matching items that describe what customers expect from an excellent company. Respondents were asked to indicate their perceptions and expectations level for each of the 52-items in the questionnaire using a seven-point Likert scale (ranging from 1 indicating ‘very strongly disagree’ to 7 indicating ‘very strongly agree’). Most scaling techniques used in marketing research use ratings on linear interval scale such as the
seven point scale popularly used to measure service quality (Kim et al., 2004). Likert scales are commonly used when measuring perception level. It is defined as a scale "which consists of a number of evaluative statements concerning an attitude object" (Dillon et al., 1994).

Section B deals with the use of a single-item direct measures of overall service quality, namely ‘overall quality of LIC is excellent’ at seven-point Likert scale. Section C deals with demographics characteristics of customers’ viz., gender, age, marital status, education, place of residence, occupation, monthly income, total number of policies bought, and mode of premium payment.

Only one set of questionnaire has been used for making an assessment of SERVQUAL instrument and to measure customers’ perceived service quality. The analysis of survey instrument has been carried out into two stages. In the first stage, analysis was done for the 16 statements (modified version of the Parasuraman et al., 1985, 1988 instrument) with the help of reliability and factor analysis (see chapter VII). In the second stage, the total 52 statements were (modified version of the Sureshchandar et al., 2001 instrument) analyzed to extract the relevant factors important for the LIC (see chapter VIII).

3.1.1.3 Construction of Questionnaire for Agents

In order to study the perception of agents, one set of a well structured questionnaire has been prepared namely the questionnaire for agents. The agents’ questionnaire (Annexure II) was divided into two sections. A pool of 39 statements/items was developed on a seven-point Likert scale ranging from ‘very strongly agree’ (1) to ‘very strongly disagree’ (7), to bring out the factors influencing agents’ perception towards LIC. Section A deals with the statements that describe the agents’ perception towards LIC. Section B deals with demographic characteristics of agents’ viz., gender, age, marital status, place of residence, education, monthly income, nature of job, occupation, type of agents, experience, club membership, and number of policies sold.

Each questionnaire was accompanied by a covering letter, which stated the purpose of research, and encouraged voluntary participation by the respondents, besides, ensuring the anonymity of their responses.
3.2 Data Analytic Methods

Commensurate with the objectives of the study, data so collected was tabulated and analysed by using appropriate statistical tools and techniques. All quantitative data was entered in a Statistical Package for Social Software (SPSS) 11.5 and analysed using — Item and reliability analysis, Factor analyses, Descriptive statistics (frequencies and percentages etc.), Weighted average scores, Correlation analysis, Multiple regression analysis, and ANOVA-one way analysis. Besides, secondary data has been analysed by computing the Growth rate and the Compound growth rate of different years.

The above said statistical measures are as follows:

3.2.1 Item and Reliability Analysis

An instrument is considered reliable when the outcome of the measuring process is reproducible. According to Zikmund (2000, p 280), reliability can be defined as “the degree to which measures are free from error and therefore yield consistent results”. Hair et al. (2010) defined reliability as an assessment of the degree of consistency between multiple measurements of a variable. According to the suggestions made by Churchill (1979), an iterative-scale purification procedure was used to develop a reduced, more parsimonious scale. Further, Churchill (1979) suggested the need to purify measures at an early stage of research. When factor analysis is conducted before purification, there is a tendency to produce many more dimensions than can be conceptually identified. So, before applying factor analysis, item and reliability analysis was performed to retain and delete scale items for the purpose of developing reliable scale. It may be pointed out that numerous marketing and organization behaviour-based-scale development articles illustrate the use of items and reliability analysis for trimming and retaining items (Netemeyer et al., 2003, p 165). Reliability of the scale shows the extent to which a scale produces consistent results, if measurements are repeated. It can be defined as the extent to which measures are free from random error (Malhotra, 2007, p 314).

For item and reliability analysis, corrected-item-to-total correlations and Cronbach’s alpha statistics were employed in the present study. Corrected item-to-total correlations reflect the extent to which any one item is correlated with the remaining items in a set.
of items under consideration. Items with low corrected item-to-total correlations are candidates for deletion (Netemeyer et al., 2003). If item to total correlation is high, then it is assumed that the item is valid. If item to total correlation is low, then we could drop that item from the scale (Nargundkar, 2008, p 64). Kerlinger (1978) recommended minimum item-to-total correlations of 0.20 or above for inclusion of items in a scale whereas Bearden and Netemeyer (1998) advocated corrected item-to-total correlations of 0.35 or above. In the views of Churchill (1979), items with correlations near zero would be eliminated from the scale. In addition, these values can be used as references to remove weak items from the scales (Lai et al., 2005).

Scale reliability would be tested by calculating Cronbach's alphas for the several factors per construct class. Cronbach's alpha is a measure of internal consistency, reliability and assesses the homogeneity of a set of items. Technically, this comes down for determining the proportion of variance in a measurement scale that is systematic (Nunally and Bernstein, 1994). Cronbach’s alpha test of reliability is the most popular estimate for measuring the internal consistency (reliability) of items in a scale, in other words, it measures the extent to which the responses collected for given item correlate highly with each other. The higher the reliability score is, the more reliable the measured construct is. Finally, what is important to be considered is that the scores increase when the number of items in a scale increases (Garson, 2002). Applying this test specifies whether the items pertaining to each dimension are internally consistent and whether they can be used to measure the same construct (dimension).

Cronbach’s alpha coefficient varies from 0 to 1, but satisfactory value is required to be more than 0.7 for the scale to be reliable (Malhotra, 2007, p 315; Nargundkar, 2008, p 64; Hair et al., 2010, p 91; Nunnally and Bernstein, 1994). The alpha coefficients for each of the dimensions should be well above the modest and satisfactory range of 0.50 to 0.60 recommended for early stages of research (Nunnally, 1979). It is worth mentioning here that an alpha coefficient of 0.6 and above is considered to be good for research in social sciences (Cronbach, 1990). With the above criteria in mind, scale refinement and data purification has been employed through item and reliability analysis.
3.2.2 Factor Analytic Technique

The main purpose of factor analysis is data reduction while retaining the underlying structure of the data set. In this way a new set of variables is created, much smaller in number, which can be used for subsequent multivariate analysis. Factor analysis only makes sense when some underlying structure does exist in the set of variables, i.e. there must be sufficient high correlations to justify the application of factor analysis (Tabachnick and Fidell, 2001). The objective of factor analysis is to find a way of condensing the information contained in a number of original variables into a smaller set of variables (factors) with a minimal loss of information (Hair et al., 2010, p 17).

Choice of the method used to extract factors is an important step in factor analysis. Principal components analysis is recommended when the primary concern is to determine the minimum number of factors that will account for maximum variance in the data for use in subsequent multivariate analysis (Malhotra, 2007, p 646). According to Hinkin (1995); and Chatfield and Collins (1980), principal components analysis is the most frequently reported factoring method for reduction of the dimensionality of the data set and formulation of new meaningful variables to describe the problem.

In order to analyse the collected data and confirm the usefulness of the proposed instrument in the context of life insurance sector, factor analysis on the scale items with the principal component analysis as an extraction method and varimax as rotation method with Kaiser Normalization has been used.

However, the adequacy or appropriateness of data for factor analysis has been examined beforehand with the help of following measures:

- **Correlation Matrix**: A correlation matrix is lower triangle matrix showing the simple correlations, \((r)\), between all possible pairs of variables included in the analysis. The diagonal elements, which are all 1, are usually omitted. A visual inspection of correlation data matrix can reveal whether there are enough correlations to go ahead with factor analysis or not.

- **Anti-image Matrix**: Matrix of the partial correlations among variables after factor analysis, representing the degree to which the factors explain each other in the
results. The diagonal contains the measures of sampling adequacy for each variable, and the off-diagonal values are partial correlations among variables (Hair et al., 2010, p 91). For the true factors to exist in the data, the values should be greater than 0.5 (Field, 2000; www.utexas.edu).

- **Kaiser-Meyer-Oklin (KMO) Measure of Sampling Adequacy (MSA):** It is another useful method to measure, to quantify the degree of inter-correlations among the variables and the adequacy of data for factor analysis. It represents the ratio of the squared correlation between variables to the squared partial correlation between variables and measures the extent to which variables belong to each other. The index ranges from 0 to 1 (Kaiser, 1974). High values (between 0.5 and 1.00) indicate that factor analysis is appropriate and values below 0.5 imply that factor analysis may not be appropriate (Malhotra, 2007, p 642). Kaiser characterises measures in the 0.90’s as marvellous, in the 0.80’s as meritorious, in the 0.70’s as middling, in the 0.60’s as mediocre, in the 0.50’s as miserable and below 0.50 as unacceptable (Norusis, 1994, p 53).

- **Bartlett’s test of Sphericity:** It is other statistical test applied in the present study for verifying the appropriateness of data set for factor analysis. This widely used method test the null hypothesis that the original correlation matrix is an identity matrix. If the correlation matrix resembles an identity then all correlation coefficients would be zero. However, for factor analysis there must be significant correlation among at least some of the variables. The researchers should note, however, that increasing the sample size could cause Bartlett’s test to become more sensitive to detect correlations among the variables (Bartlett, 1950). Bartlett’s test of sphericity reached the threshold value of .05 for significance (p = .000) (Hair et al., 2010).

Several criteria guided the decision regarding how many factors should be extracted from the exploratory factor analysis. These are:

- **Determination based on Communality Values:** Communality is used to measure the percentage of variance in a given variable explained by all the factors together. It can be interpreted as the reliability of the indicator. If the communality value of given variable is low, then this variable should be probably removed from the model,
because the factor it pertains to cannot explain its variance enough. Nevertheless, the interpretation of the values of communalities should be done in relation to the interpretation of the factors. The table on Communalities consists of ‘initial and extracted’ values. The initial values will be always 1.0 as in this case the number of factors is equal to the number of variables. The extracted value represents the percentage of variance in a given variable explained by the extracted factor. As in this case, the extracted value will be lower than 1.0 (Garson, 2002). The factor solution should explain at least half of each original variable's variance, so the communality value for each variable should be 0.50 or higher (Field, 2000; Hair et al., 2010, p 118; www.uic.edu).

- **Determination based on Eigen Values:** Eigenvalue represents the total variance explained by each factor (Malhotra, 2007, pp 642, 647). During factor analysis, factors with eigenvalue of more than 1 signify that the variables met the acceptable standard of validity analysis and would be retained for further analysis (Hair et al., 2010, pp 108, 110; Pallant, 2005). If the eigen value of a given factor is low, this means that this factor explains little of the variance in the variables and can be dismissed from the model. Thus, the eigen value measures the amount of variation in the variables which is explained by a given factor (total variance explained should be equal to the number of variables) (Garson, 2002).

- **Determination based on Scree Plot:** Another approach for determining the number of factors to be extracted is by using the Scree plot (Catell’s scree test). According to this method, the number of factors to be extracted should equal the number of points above the point at which the shape of the curve on the Scree plot changes its direction and becomes more horizontal (Pallant, 2005).

- **Determination based on Percentage of Variance:** This is the percentage of the total variance attributed to each factor. However, it is recommended that at least 60 percent of the variance in the data had to be explained with the final factor solution (Malhotra, 2007, pp 642, 647).

- **Determination based on Factor Loadings:** Factor loadings are simple correlations between the variables and the factors (Malhotra, 2007, p 642). The factors having loadings greater than 0.40 are considered important for retention purpose (Dixon,
Data Base and Methodology

1997; Gray-Donald *et al.*, 2007). Factor loadings greater than ± 0.30 are considered to meet the minimum level; loadings of ± 0.40 are considered more important; and if the loadings are ± 0.50 or greater, they are considered practically significant. A factor loadings of ± 0.30 is regarded as significant for a sample size of 350 or greater (Hair *et al.*, 2010, p 116). In this study loading of ±0.50 has been used to retain variables. In case of customers’ data, cut-off point used is ±0.50 as the sample size is less than 350. In case of agents’ data, cut-off point used is ±0.50 as the sample size is less than 250. In order to make the output of the analysis easier to understand, Rotation can be used which is also necessary to help for the interpretation of factors. After Rotation is performed, the eigenvalues of given factors and the factor loadings will be changed. There exist number of Rotation Methods and one of the most common ones is the Varimax Rotation Method, which makes it possibly easy to associate given variable with a single factor (Garson, 2002). According to the Harrison-Walker’s (2001) study, it was likely that two or three factors should be present.

3.2.3 Weighted Average Scores (WAS)

Weighted average scores have been calculated to find out the importance attached by respondents to given variables rated by them on a seven point Likert scale. In the present study, this technique has been used to identify areas where quality improvement is needed the most.

3.2.4 Multiple Regression Analysis

Multiple regression analysis is a powerful and flexible procedure for analyzing associative relationships between a single (metric) dependent variable and one or more independent variables. Regression can be used in the following ways:

- Determine whether the independent variables explain a significant variation in the dependent variable: whether a relationship exists
- Determine how much of the variation in the dependent variable can be explained by the independent variables: strength of the relationship
- Determine the structure or form of the relationship: the mathematical equation relating the independent and dependent variables.
Data Base and Methodology

- Predict the values of the dependent variable
- Control for other independent variables when evaluating the contributions of a specific variable or set of variables (Malhotra, 2007, pp 572-573, 682)

To assess the overall effect of the instrument on service quality and to determine the relative importance of the individual dimension, this technique has been used in the present study.

3.2.5 ANOVA-One Way Analysis
ANOVA has been applied to test the null hypothesis that there is no significant difference within the various demographic groups as regards the ratings of different variables. This technique has been performed to test whether any significant differences existed among various groups of respondents (i.e. customers and agents) across their perception scales.

3.2.6 Growth Rate/Compound Growth Rate
Based on the empirical research (Garg, 2007; Arora, 1992; Bawa, 2006; and Gupta, 2007), the growth rate and the compound growth rate have been computed to evaluate the performance of LIC and private insurers for different years.

3.3 Limitations of the Study
The conclusion drawn from the results of this study are subject to certain limitations.
- The basic shortcoming of primary survey may be the gap between the respondents’ speculative response and the truth, as observed by Malhotra (2007, p 153) that the main disadvantage of conducting a primary survey is that respondents may be unable or unwilling to provide the desired information. In the present study too, the verity might be quite different from the stated response.
- This study has been carried out in the perimeters of Punjab and the results obtained thereof may not be pertinent to the country as a whole. However, the study could serve as a lead and be made extensive to other states of India.
- Again, the present study is restrictive in context, conducted by taking a sample of 337 customers and 225 agents of LIC (a public company), ignoring the private life
insurance companies. This cannot lead to the generalizability of the findings and the results may not be implied conclusively to the whole life insurance industry.

- In the case of marketing strategies, ‘socio-economic profile’ published at divisional level (the relevant source of information regarding agents, development officers’ strength and their productivity etc.) has ceased to be published by the respective divisions. Hence, the non-availability of the relevant information has put a few constraints on the study.

These limitations may lessen the ability of generalizing the results of this study to other life insurance companies’ settings. Nonetheless, the conceptual and methodology limitations of this study need to be considered when designing future research.