CHAPTER III

METHODOLOGY

The methodology adopted for the present research is explained in detail in this chapter. The key components of methodology namely, research design of the study, the variables examined, hypotheses formulated, universe, sample and sampling procedure, tool for data collection, data processing and data analysis have been explicitly described. The chapter ends with a detailed description of the statistical tools that were applied to analyse the primary data.

3.1 The Research Design

As the primary aim of the present research is to determine the relationship between the independent variables (Eg. social, economic and demographic characteristics of the respondents) and the dependent variable (Eg. the extent and forms of cybercrime victimisation) in a population, it broadly falls under quantitative research. According to Singh (2007), a quantitative research design can be divided into two types namely, exploratory research and conclusive research. He states that when the scope of the research is not clear or the alternative options have not been clearly defined, one may choose exploratory research. This method of research allows the researcher to explore research issues in detail. Exploratory research
also helps the researcher to familiarise himself to the problem or concept to be studied. It can also be called as initial research, since it forms the basis for a more conclusive research. Along the same lines, Babbie (2004) explained the purpose of exploratory research. According to him, exploratory studies are mostly conducted for three purposes: (i) to satisfy the researcher’s curiosity and desire for a better understanding of the research issues; (ii) to test the feasibility of undertaking a more extensive study; and (iii) to develop the method to be employed in any subsequent studies. In view of the above criteria, the research design followed in the present study is the exploratory research design.

3.2 Variables Investigated

1) Social, economic and demographic characteristics of respondents such as: Age, gender, nativity, current place of stay, education and approximate monthly family income

2) Number of computers owned, Internet connectivity, duration of Internet usage, place of Internet usage, purpose and frequency of Internet usage

3) Extent of computer knowledge

4) Extent and forms of cybercrime victimisation

5) Impact of cybercrime victimisation
6) Reporting behaviour of respondents

7) Needs and expectations of respondents

3.3 Hypotheses

1) There is no significant difference among the mean ranks of the forms of cybercrime victimisation.

2) There is no significant association between the age, gender, nativity, current place of stay, course of study, monthly family income of respondents and cybercrime victimisation.

3) There is no significant difference between the age, gender and nativity, current place of stay, course of study and monthly family income of respondents and the extent and forms of cybercrime victimisation.

4) There is no significant association between the number of computers owned, Internet connectivity, duration of Internet usage, place of Internet usage, purpose and frequency of Internet usage and the cybercrime victimisation.

5) There is no significant association between the social, economic and demographic characteristics of respondents and the extent of cybercrime.

6) The impact of cybercrime victimisation varies according to the forms of cybercrime.
7) There is no significant association between the persons to whom cybercrime victimisation was reported (reporting behaviour) and the extent of cybercrime victimisation.

3.4 Universe, Sample and Sampling Procedure

3.4.1 Locale of the Study

It may be pertinent to briefly explain why the researcher has chosen Chennai city as the locale of the present research. Apart from the reason that the researcher resides in Chennai, it is the capital city of the state of Tamilnadu and is a major educational, cultural and economic centre. The city is also host to the sixth largest population in India with a population of 46.47 lakhs according to the Census 2011. It is also relevant to mention that the average literacy rate in Chennai city is 90.2 percent with the male literacy rate at 93.7 percent and female literacy rate at 86.6 percent (Directorate of Census Operations, 2011). According to the Tamilnadu Directorate of Collegiate Education (TNDCE, n.d.), there are 74 Government Arts and Science colleges, 37 University constituent colleges, 139 aided Arts and Science colleges and 443 self-financing Arts and Science colleges making it a total of 693 Arts and Science colleges in Tamilnadu. University of Madras, Chennai has 132 Arts and Science colleges (including Government, Aided and Self-financing colleges) in its fold, which constitutes about 19 percent of the Arts and Science colleges in Tamilnadu (College Development Council, University of
Madras, 2014). According to the list of colleges and details from TNDCE, there are 54 colleges located within the city, which constitutes 41 percent of the colleges that are affiliated to the University of Madras.

The researcher has chosen college students as respondents for the present study as many studies have been carried out on cybercrime victimisation among school students in various parts of the world including India but only a very few studies on cybercrime victimisation have been carried out among college students. Therefore, the present study would open up several new possibilities of research on various dimensions of cybercrime victimisation in the future.

### 3.4.2 Universe of the Study

The universe of the present study is the student community from the various Arts and Science colleges affiliated to the University of Madras. Of the 132 Arts and Science colleges, 54 colleges that are situated within Chennai were identified. Therefore, the population of the study are under-graduate and post-graduate students studying various arts and science courses such as B.A., B.Sc., B.Com., B.C.A., M.A., M.Sc., M.Com., M.C.A. and M.B.A..
3.4.3 Sample and Sampling Procedure

The 54 colleges were broadly classified into co-education colleges (16), colleges for women (24) and colleges for only men (14). The colleges were listed based on these three categories and using a systematic random selection method; every third college was identified, thereby totalling 18 colleges. As the researcher could not prepare a complete list of students studying in the various courses, it was decided to adopt the purposive sampling method. However, while selecting the sample, the researcher kept in mind certain crucial factors. That is, the researcher wanted to select a sample so as to have a fair representation of the students studying in each college. In other words, while identifying the respondents for the study, the researcher considered factors including gender of the respondents and the course and year of study. Through this process a total sample of approximately 1,800 students were approached and requested to participate in the study. Of them, 1,600 participated and completed the questionnaire. Of 1,600 questionnaires, 80 questionnaires were found incomplete and hence were disregarded. Thus, for the present research the total sample size was 1,520. The profile of the respondents has been presented in the chapter on Results and Discussion (Chapter IV).
3.5 Research Tool

3.5.1 Construction of the Tool

Keeping the objectives of the study in mind and also taking into account the nature of the respondents, the researcher decided to construct a structured and closed-ended questionnaire. To measure all the variables of the study, a list of questions was drafted. The preliminary draft of the questionnaire had nearly 30 main questions and was first vetted by the supervisor, co-researchers and a few post-graduate students from the Department of Criminology, University of Madras. Relevant changes were made to the questionnaire based on their feedback. The researcher's many meetings with the guide and supervisor helped in drafting questions/items that would cover all the objectives of the research. The frequent discussions also helped ensure the questions were coherent, easily understood and logically ordered. Items such as using ‘my old sim card to make calls to me’ ‘sharing/posting anti-religious comment on my Facebook/blogs’ and ‘stealing my personal data from my mobile/sim and sending it to others’ were included under the main question relating to the frequency of cybercrime victimisation experienced, which was meant to measure the variable ‘extent and forms of cybercrime victimisation’. However, after discussions with the guide and co-researchers, these questions were omitted. The researcher also conducted a group discussion with students pursuing their post-
graduate degree at the Department of Criminology of the University of Madras on the topic of the research using the questionnaire. The participants of the group discussion offered their comments and suggestions to further improve upon the questionnaire. Following the group discussion and initial feedback, the fourth part of the questionnaire on victimisation comprised 29 items.

Having arrived at a definitive number of logical and meaningful questions, the questionnaire was then assessed by a professor of English for language correction and also a statistician for correcting any obvious statistical errors that would create problems during analysis. Their suggestions were also incorporated into the questionnaire before arriving at the revised questionnaire. The questionnaire was not translated into Tamil as the terms used in the questionnaire were technical in nature and translating it into Tamil was tedious and cumbersome. However, due care was taken not to explicitly mention the phrase ‘cybercrime victimisation’ in the questionnaire, as it was likely to put off several students and not elicit true responses.

3.5.2 Pilot Study

From the population of the study, a sample of 30 respondents were chosen and the questionnaire was administered to them. The data collected, processed and analysed using the Statistical Package for the Social Sciences (SPSS). To test both the reliability and validity of
the tool, Cronbach’s alpha coefficient was applied. Cronbach’s alpha co-efficient is a measure of internal consistency or how closely related a set of items are as a group. It was found that Cronbach’s alpha score was 0.853 for the variable ‘purpose of using the Internet’, 0.845 for ‘extent and forms of cybercrime victimisation’ and 0.892 for the variable ‘impact of victimisation’. All these values generally desired standard of value 0.7 or higher. Therefore, the items under the variables indicated high internal consistency.

3.5.3. Description of the Tool

The final questionnaire was divided into seven parts with a total of 22 main questions.

The content of the questionnaire is as follows:

- Part I: Profile of the respondents – comprises six items such as age, gender, nativity, current place of stay, course of study and monthly family income.
- Part II: Includes nine items relating to the duration and purpose of Internet usage.
- Part III: The questions under this part measure the extent of the respondents’ knowledge of computers and its usage.
- Part IV: Comprises 29 statements with a five-point scale to measure the extent and forms of victimisation.
Part V: Includes questions to measure the impact of the cybercrime victimisation.

Part VI: Questions 19–21 measure the reporting behaviour of the respondents.

Part VII: Comprises of one question to measure the needs and expectations of the respondents (refer Appendix A).

3.6 Method of Data Collection

The questionnaire is in English, using simple and clear language making it easily self-administered. After getting the necessary prior permission from college officials (in some colleges with a written letter from the researcher himself and in some other places, an official letter from the Guide), the researcher met the respondents in groups mostly in a classroom set up. At the beginning of the session, the students were briefed on the topic and only Internet users were requested to take up the questionnaire. Following the completion of questionnaires with the group, the researcher spent time with each group discussing the problems of cybercrime victimisation. During these discussions, the researcher found that the respondents had several doubts and clarifications on the problem of cybercrime victimisation. The respondents were able to highlight these in the section dealing with the needs and expectations of the respondents. To maintain ethicality, the researcher included a note at the beginning of the questionnaire to assure respondents that the information
provided by them would be treated as confidential. At the end of the questionnaire, the e-mail address and the official address of the researcher is mentioned for queries, if any.

3.7 Data Processing and Data Analysis

3.7.1 Processing of the Data

The data collected from 1,520 respondents was processed before it was subjected to statistical analysis. As stated, key variables were measured on four point and five point scales. The responses obtained were assigned appropriate scores.

- The various purposes and the frequency of Internet usage are measured on a four point scale as follows: Very often = 4, Often = 3, Sometimes = 2, Not at all = 1. Higher the score, higher the frequency of Internet usage.
- Similarly, the extent of computer knowledge is measured using a scale ranging from ‘to a large extent’ to ‘not at all’. For the response ‘to a large extent’ the score assigned is 4 and for the response ‘not at all’ the score assigned is 1.
- With regard to the extent and forms of cybercrime victimisation, the responses are measured on a five point scale namely – very often, often, sometimes, rarely and never and the scores assigned are 5, 4, 3, 2 and 1 respectively. Higher the score, higher the extent of victimisation faced by the respondent. It is to be noted
that cybercrime victimisation is not treated as one entity since it includes several types or patterns of victimisation. These types and patterns of victimisation are arrived at using factor analysis techniques, explained below.

- Impact of victimisation is also measured on a five point scale ranging from ‘very often’ to ‘never’. ‘Very often’ is assigned a score 1. If a respondent has obtained a higher score it means that the impact of cybercrime victimisation on him/her is high.

3.7.2 Factor Analysis

To measure the forms of cybercrime victimisation, 29 items were included in the questionnaire. Factor analysis was carried out to reduce this lengthy number of items into a few factors (forms of victimisation).

There are two types of factor analysis:

**Exploratory Factor Analysis (EFA)**

- It is applied when there is no predefined idea of structure or the number of dimensions in a set of variables.

**Confirmatory Factor Analysis (CFA)**

- It is applied when a specific hypothesis about the structure or the number of dimensions underlying a set of variables is to be confirmed.
According to Byrne (2010), *a priori* knowledge of the researcher - a good understanding and knowledge of theory, empirical research or both - in designing the questionnaire for the survey may have some hypotheses formulated as a relationship exists between observed variables and their underlying constructs. This has to be statistically tested for further analysis of the research problem. CFA helps the researcher to statistically test the hypothesis. CFA uses a path analysis diagram to represent variables and factors. Structural Equation Modelling (SEM) is typically used for performing CFA. In this study, the researcher has used only confirmatory factor analysis to confirm the *a priori* factors among the cybercrime victimisation items.

The Cronbach’s alpha for various dimensions of cybercrime victimisation is as follows: Cyber harassment = 0.776; Cyber stalking = 0.756; Cyber hacking = 0.791; and Cyber sexual victimisation = 0.768. These results indicate good internal consistency and so, no deletion from the scale was needed as the item correlations and alpha reliabilities for all sub-scales were relatively high. However, one item from one of the forms of victimisation was deleted.

To proceed with measuring the fit in CFA, some interrelated statistical techniques were used to analyse the data as a supportive stream. This process follows the development of an individual measurement model for each construct measure to CFA and the
overall measurement model to check the dimensionality of the construct and validity of the measures.

Structural Equation Modelling (SEM) is a statistical methodology that takes a confirmatory (i.e., hypothesis-testing) approach to the analysis of a structural theory bearing on some phenomenon (Byrne, 2010). SEM is a quantitative data analytical technique, which specifies, estimates and tests theoretical relationships between observed endogenous variables and latent, unobserved exogenous variables. SEM does not designate a single statistical technique but rather a family of relevant procedures including analysis of covariance structure and this combines regression and factor analysis as well. The SEM approach starts with model specification that links the variables assumed to affect other variables and directionalities of those effects (Kline, 2011). In the present study, to examine the fit on the constructs, a maximum-likelihood estimation of CFA in Analysis of Moment Structures (AMOS version 21.0) was conducted.

First, fit statistics should be evaluated to check whether the proposed model is fit to the data or not, or whether any modification is required to increase fit. Though there are various fit indices to measure the Goodness-of-Fit statistics, the present study uses a set of some of the indices to check the fitness of the model.

In the present study, there are four hypothesised factors namely, cyber harassment, cyber stalking, cyber hacking and cyber sexual
victimisation. Since it is not needed to find all sorts of fit measures in a report, however, a subset or sample of fit indices from major categories has been reported in this study to assess the degree of overall fitness of the measurement model and the structural model. Taking sample sensitivity and model complexity effect into account, $\chi^2$/df (CMIN/DF), p value, Goodness-of-Fit Index (GFI), Adjusted Goodness-of-Fit Index (AGFI), Comparative Fit Index (CFI), Root Mean Residual (RMR) and Root Mean Square Error Approximation (RMSEA) were considered in this study for evaluating fit indices because these indices have been commonly used. The following table shows how various indices have resulted in the test.

Table 3.1: Fit indices on the forms of cybercrime victimisation

<table>
<thead>
<tr>
<th>Factors</th>
<th>Cyber harassment</th>
<th>Cyber stalking</th>
<th>Cyber hacking</th>
<th>Cyber sexual victimisation</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of items before CFA</td>
<td>8</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Chi-square value</td>
<td>8.054</td>
<td>14.374</td>
<td>13.048</td>
<td>14.592</td>
</tr>
<tr>
<td>P value</td>
<td>0.328</td>
<td>0.073</td>
<td>0.071</td>
<td>0.068</td>
</tr>
<tr>
<td>GFI</td>
<td>1.000</td>
<td>0.997</td>
<td>0.998</td>
<td>0.997</td>
</tr>
<tr>
<td>AGFI</td>
<td>0.994</td>
<td>0.991</td>
<td>0.990</td>
<td>0.991</td>
</tr>
<tr>
<td>CFI</td>
<td>1.000</td>
<td>0.997</td>
<td>0.998</td>
<td>0.997</td>
</tr>
<tr>
<td>RMR</td>
<td>0.007</td>
<td>0.011</td>
<td>0.008</td>
<td>0.009</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.010</td>
<td>0.023</td>
<td>0.024</td>
<td>0.023</td>
</tr>
<tr>
<td>No. of items after CFA</td>
<td>7</td>
<td>7</td>
<td>7</td>
<td>7</td>
</tr>
<tr>
<td>Cronbach’s alpha</td>
<td>0.776</td>
<td>0.756</td>
<td>0.791</td>
<td>0.768</td>
</tr>
</tbody>
</table>

Analysing the indices, Chi-square value is the traditional measure for evaluating overall model fit in covariance structure models and it
assesses the magnitude of discrepancy between the sample and fitted covariance. However, it is said that a good model fit would show a significant result of \( p > 0.05 \). The Chi-square value is often referred to as either a goodness-of-fit or badness-of-fit measure, whereby large values correspond to bad fit and small values correspond to good fit, but they are often referred to as problematic (Kline, 2011). The degree of freedom serves as a standard by which to access Chi-square is small or big. Taking into account these complexities some researchers have referred to and preferred the use of ‘normed’ Chi-square measure that is measured by the equation CMIN/DF. Byrne (2010) suggested that if the value of normed Chi-square was greater than 1 and smaller than 2, then it would indicate a very good model fit. In all the four factors of the present study, CMIN/DF ratio (Factor I: 1.151, Factor II: 1.797, Factor III: 1.864 and Factor IV: 1.824), which indicated a good model fit. The other model fits such as GFI, AGFI, CFI and RMSEA were also employed to access the goodness of fit of the measurement model.

To improve the relative improvement in fit to the model, the researcher measured other values such as Comparative Fit Index (CFI). It is a revised form of the Normed Fit Index (NFI), which takes into account sample size (Byrne, 2010) that performs well when the sample size is small. This statistic assumed that all latent variables were uncorrelated (null/independence model) and compared the sample covariance matrix with this null model (Hooper, Coughlan &
Mullen, 2008). They also recommended that a value of CFI greater than or equal to 0.95 was indicative of good fit. The CFI values of four factors in the present study were 1.000, 0.997, 0.998, and 0.997 respectively, which is according to the acceptable norms, were indicative of good fit.

Another fit index that is used as an alternative to the Chi-square test is Goodness Fit Index (GFI). It would calculate the proportion of variance that would be accounted for by the estimated population covariance. The statistics range from 0 to 1 with larger samples increasing its value. Traditionally, an omnibus cut-off point of 0.90 had been recommended for the GFI; however simulation studies showed that when factor loadings and sample sizes were low a higher cut-off of 0.95 would be more appropriate (Shevlin, Miles & Lewis, 2000). The present study showed that the four factors I, II, III, and IV have GFI values of 1.000, 0.997, 0.998 and 0.997 respectively and so it was also indicative of good fit model. Related to the GFI is the AGFI, which adjusts the GFI based upon degrees of freedom with more saturated models reducing fit. AGFI tends to increase with sample size. As with the GFI, values for the AGFI also range between 0 and 1 and it is generally accepted that values of 0.90 or greater indicate well-fitting models (Hooper et al., 2008). The present study showed the AGFI values as 0.994, 0.991, 0.990 and 0.991 for the four factors, which indicated a good fit model.
Root Mean Residual (RMR) and Root Mean Square Error of Approximation (RMSEA) are two other measures of fit, considered in this present study. RMR is the square root of the difference between the residuals of the sample covariance matrix and the hypothesised covariance model (Hooper et al., 2008). RMSEA helps to find how well the model, with unknown but optimally chosen parameter estimates would fit the population covariance matrix (Byrne, 2010). Hu and Bentler (1998) suggested that RMR values to be closer to 0.08 or below and RMSEA values to be closer to 0.06 or below for good model fit. The present study also found that the RMR values for the four factors to be 0.007, 0.011, 0.008 and 0.009 and RMSEA values to be 0.010, 0.023, 0.024 and 0.023, indicating the model fit.

After analysing the estimates and checking the goodness of fit of variables, the researcher removed only one item (Receiving threatening mails) in Factor I, making the number of items for each factor to be 7. However, the other correlated items were kept intact so that the researcher could analyse the repeated measures even if they are correlated. Absolute fit indices determine how well the model fits the sample data and which model represents the superior fit (Hooper et al., 2008). Based on the overall Goodness-of-Fit statistics, the four-factor model for cybercrime victimisation yielded perfect fit statistics after removing the item with the lowest co-efficient value (refer Appendix II for item-total statistics of each factor).
The path diagrammatic representation of the estimates is illustrated below. Before removing the correlated items, the path diagram of Factor I is presented below (Figure 3.1). The numbers in the rectangular boxes are the item numbers in the questionnaire.

Figure 3.1: The path diagram of cyber harassment (CH) before correction

The path diagram of Factor I after removing the item ‘Receiving threatening e-mails’ is as follows:
Figure 3.2: The path diagram of cyber harassment (CH) after correction

The path diagram of Factor II is as follows:

Figure 3.3: The path diagram of cyber stalking (CS)
The path diagram of Factor III is as follows:

**Figure 3.4: The path diagram of cyber hacking (CHA)**

The path diagram of Factor IV is as follows:

**Figure 3.5: The path diagram of cyber sexual victimisation (CSV)**
3.7.3 Application of Statistics

Besides the factor analysis, both descriptive and inferential statistics have been applied to analyse the data of the present study.

- **Descriptive Statistics:** Frequency and percentages have been used and the results obtained thereof are presented in the form of tables and diagrams.

- **Inferential Statistics:** To test the hypotheses formulated for the present study, the following bi-variate and multi-variate techniques have been applied.
  
  - Friedman Test was applied to test the significant difference among the mean ranks of forms of cybercrime victimisation, among mean ranks of various purposes of Internet usage and among mean ranks of various operations of knowledge in computer.
  
  - Chi-square was used to test the association between variables such as age group, gender, locality, current place of stay, course of study, monthly family income, number of computers at home, number of hours connected to the Internet per day, and accesses of the Internet at home, friend’s place, net café, and college, computer (desktop/laptop) at home and cybercrime victimisation. It was also used to test the association between gender, current place of stay, the number of hours connected to the Internet and the persons to whom cybercrime victimisation was reported and levels of cybercrime victimisation.
• Analysis of variance (ANOVA) was applied to establish the significant difference between age group, locality, current place of stay, monthly family income and the number of hours connected to the Internet per day and the extent and forms of cybercrime victimisation.

• t-test was used to test the significant differences between gender, course of study and Internet connection at home and the extent and forms of cybercrime victimisation.

• Pearson’s Correlation Co-efficient was used to find the relationship between the extent and forms of cybercrime victimisation and the extent of knowledge in computer.

• Regression Analysis was applied to determine the statistical relationship between the various forms of cybercrime victimisation and the impact of cybercrime victimisation.

• Classification Tree was used to simplify and split the data into small branches. It was used to identify the relationship between age, gender, locality, year of study, current place of stay, number of computers at home and number of hours connected to the Internet per day and overall cybercrime victimisation.