CHAPTER 3: RESEARCH METHODOLOGY

Data and methodology are some of the important elements for a successful survey research. This chapter reviews the sampling, variables and analysis techniques for the project.

3.1   Sampling Procedure

Systematical planning during the preliminary stage of survey can help ensure the success of the project. In this case, sampling is one of the important elements for survey. The plan for sampling needs to include the necessary precision in terms of a bound on the error of estimation with the given cost and time period. For these, there are several factors to be considered in the planning stage as the following:

- Research design
- Target population
- Sampling methodology
- Survey selection criteria

These factors need to be clearly defined, reviewed and understood to avoid uncertainty during fieldwork and analysis. By doing these, the possibility of getting the wrong target audience can be minimized, and that the selected respondent would represent the population and reflect the true objective of the study.

For this project, probability sampling is used to access various characteristics of target population. This type of sampling is preferred because it permits statistical inferences of the population. The measurements or observations made on the sampling units are called variables and their values vary over different instances. In this case, most of the data collected are discrete variable with ordinal values. Where, it can only assume specific integer values. The reason is due to the nature of "close ended" questionnaires that was designed for the survey. For each question asked, there will be a certain explanation attached to the integer for later stage of analysis.

The data set consists of general hand phone user's perception regarding the specified mobile phone unit. The information was collected previously by ACNielsen (M) Sdn. Bhd., through one of its nationwide survey on the related matter.

The survey was conducted through personal, face- to- face household interview on both male and female, aged 15 and above, of all races. Most of the coverage is in urban town of Peninsular Malaysia with emphasis at Klang Valley, Pulau Pinang and Johor Bharu. The screening criteria were that the respondents have to be hand phone user, which can be further categorized into two main groups: pre- paid users and post- paid users. The following proposed sample distributions are based on the judgmental need or preference of the client to investigate the market situation (Table 3.1). As this is only a proposed sampling plan, the actual number may vary during the fieldwork:
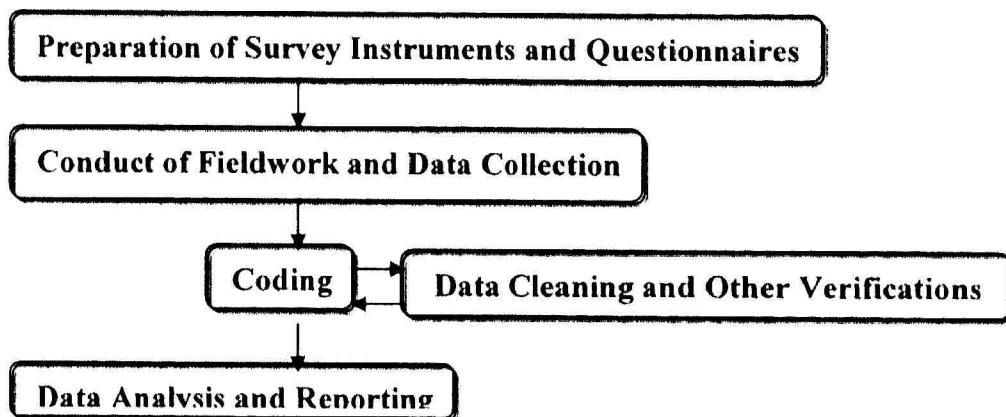
Table 3.1: Proposed Sample Distribution

Demographic	Brand	Nokia	Motorola	Ericsson	Samsung	Alcatel	Grand Total
Ethinic Groups	Malay	300	200	150	80	80	810
	Chinese	300	200	150	80	80	810
	Indian	200	100	100	40	40	480
Gender	Male	560	350	280	140	140	1470
	Female	240	150	120	60	60	630
Location	Klang Valley	200	125	100	50	50	525
	Pulau Pinang	160	100	80	40	40	420
	Johor Baru	160	100	80	40	40	420
	Central Region	80	50	40	20	20	210
	Northern Region	80	50	40	20	20	210
	Southern Region	80	50	40	20	20	210
	Eastern Region	40	25	20	10	10	105
Payment	Post paid	700	400	300	150	150	1700
	Pre Paid	100	100	100	50	50	400
Grand Total		800	500	400	200	200	2100

Five brands of telecommunication unit provider are the main focus of this survey: Nokia, Motorola, Ericsson, Samsung and Alcatel. . The reason being is that they are the most competitive telecommunication unit providers in the market and their creativity, as well as innovative approach in the market may prove to be interestingly dynamic.

As the proposed sample size has been properly identified, the next step is to prepare for data collection. The following diagram summarizes the process for this survey:

Figure 3.1: Data Collection Process

```
┌──────────────────────────────────────────────────────┐
│  Preparation of Survey Instruments and Questionnaires  │
└──────────────────────────────────────────────────────┘
                         │
                         ▼
┌──────────────────────────────────────────────┐
│  Conduct of Fieldwork and Data Collection      │
└──────────────────────────────────────────────┘
                 │
                 ▼
        ┌──────────┐        ┌──────────────────────────────────────┐
        │  Coding  │───────▶│  Data Cleaning and Other Verifications │
        └──────────┘◀───────└──────────────────────────────────────┘
             │
             ▼
┌──────────────────────────────┐
│  Data Analysis and Reporting   │
└──────────────────────────────┘
```

Data collection is an important process that needs to have close supervision. Questionnaires, authorization letters, show cards, and other survey instruments were prepared before the Field Officer actually goes into field. The questionnaire was tested on quality for "sanity" during the pilot study. Where, it is to ensure that all the questions asked are important for the objective, no replication, and do not bore or offend the respondents. The next step is the recruitment process by the Field Officer to identify any suitable respondents. After the fieldwork, the questionnaires will be coded by the Coding Officer and undergo a series of cleaning process for verification before actual analysis begun.

3.2 Variable List

It is important that during the data collection stages, the following main sources of brand preference calculation to be collected and properly coded for analysis:

Brand Usage: Consists of five major studied brands - Nokia, Motorola, Samsung, Ericsson and Alcatel. Filtering will be done if individual analysis is needed in the subsequent analysis.

Demographic Variables: This is actually a set of variables that were collected during the survey to understand the respondent background. The information collected was race, gender, state, household income and personal income.

Network Service: This variable is related to customer usage of the telecommunication network services and is coded "1"for "Post paid" and "2" for "Pre- Paid"

Brand Awareness: This is related to prompted and unprompted recall for specific brand, or in other words, the respondents' level of awareness for the brand. Five possible values can be attributed to this variable: 1, 2, 3, 4 or 5. Where, the highest level or "top of mind" is "5"; and the lowest level or "unaware" is "1".

Product and Non- Product Benefits: This relates to users' psychology association between product attributes and the brand that they are using (Naumann & Giel, 1995). In other words, the level of thinking or consideration that occur in their mind before using a particular product.

- "Product Benefit" is denoted as "Consideration" in the study. Where the denotation for respondents that will consider using the product is "1" and "0" if the respondent will not consider using the product at all.

- As discussed earlier (in Chapter 2), "Non Product Benefits" comprise symbolic attributes or benefits, it can utilize price premium as the proxy for measurement. Six possible values can be imputated into this variable. Where, if a respondent will pay "Whatever it costs", then "Premium" = "6"; and if the respondent "Wouldn't buy it at all", then "Premium" = "1"

Suggested Brand: Relates to the "commanding strength" of the brand, or in other words, the brand's influential power in affecting users' judgment to recommend the brand to his friends. Two possible values for this variable, where "1" will be given if the brand will be suggested to other people, and "0" if otherwise.

- Preferred Brand: This can be viewed as the number one brand that the respondents always use. Three values can be assigned to this variable, where the value of "1" will be given to the respondent with the answer of "not his preferred brand", and "3" for the respondent that answer this is his most preferred brand.

- Brand Attribute: There are 26 variables that act as the brand attributes for the mobile phone brand. Among the variables are "Durable", "Trusted Brand", "Best Technology", etc. Details of the variable are listed in Appendix 1.

These attributes are brand associations that were assigned with five point scales. Where, "1" means "Does not applies at all", and "5" means "Applies strongly".

3.3   Statistical Data Analysis

Competitive position plays the important role for both company and brand as an indication of overall health in the market. Measuring such volatile environment may prove to be difficult but is quite necessary. One way of doing so is by using Statistical Package for Social Science (SPSS) to analyze the data as discussed below.

3.3.1   Preliminary Data Exploration

Generally, the first step of any analysis is to investigate the data in some concise way. This can usually be accomplished by listing each variable by summarizing the data set using tables, charts and graphs. In this case, descriptive statistics is used to show the overall sample distribution. The purpose of this is threefold:

1.   to summarize the data

2.   to detect underlying trends

3.   statistical test for variable relationship

Chi- square (χ^2) test for significant is used to examine the underlying relationship between the variables. Apart from that, direct brand comparison or brand positioning are calculated according to the preference indices as per Professor Kevin Keller's modeling procedure. Detailed calculation of the technique is shown in Appendix 2. In this case, descriptive statistics will serve as the basis for data exploration to the later stage.

3.3.2   Multivariate Analysis

Multivariate Analysis is usually applied by statisticians, defined as any statistical, mathematical or graphical approach which requires multiple dependent variables be considered simultaneously in the analysis (Brian A. Rock, 1985). It is a set of techniques used when variation in several variables has to be studied simultaneously for two or more dependent variables. Discriminant Analysis and Factor Analysis are the two multivariate techniques been use in this paper.

3.3.2.1   Discriminant Analysis

One of the objectives of this paper is to understand the market structure for different types of users. In this case, discriminant analysis is used because it is able to determine which variables discriminate between two groups. Where, the underlying idea of the analysis is to determine whether groups differ with regard to the mean of a variable, and then to use that variable to predict group membership (George H. Dunteman, 1984). In other words, the analysis is to predict group membership based on a linear combination of interval variables.

Although there are some similarities between discriminant analysis and regression analysis which involve a single criterion or dependent variable and multiple predictor or independent variables, the nature of these variables differ. In regression, the dependent variable is metric or interval scaled, whereas in discriminant analysis it is categorical. Apart from that, regression analysis is usually for model estimation, but discriminant analysis is to investigate group differences or in other words, relationship between a single categorical dependent variable and a set of metric independent variables.

The procedure begins with a set of observations with both group membership and the variable values are known. The end result of the procedure is a model that able to provide insight observation to the relationship between group membership and its variables (David W. Stockburger, 1996).

The steps involved in conducting discriminant analysis consist of prediction, determination of significance, interpretation, and validation. The first step of this process is to randomly divide the sample into two parts:

1. Analysis sample: which consist of 70 percent of the total sample, to be used for estimating the discriminant function

2. Validation sample: which consist of the other 30 percent of the sample for validating the discriminant function

When assessing the validity of the discriminant analysis, no general guidelines are available to determine how much improvement should be expected from the hit ratio in the validation sample. However, some authors have suggested that classification accuracy achieved by discriminant analysis should be at least 25 percent greater that the obtained by chance.

This project utilized discriminant analysis to determine the difference exhibits between the users in term of demographic characteristics. Where, the dependent variable for this analysis is "Network Service". While the independent variables is the "Demographic" (i.e., ethnic groups, household income, personal income, age and gender). This procedure is appropriate because of the nature of the predefined categorical groups (prepaid and postpaid users), and the ordinal scales from the independent variables. The independent variables will be tested on the means of discriminant functions in all the groups by employing stepwise estimation method. This method is used because of its capability of enabling large number of independent variables for inclusion in the function by discriminating each variable sequentially and reduces any variables that are not able to discriminate between the groups.

The discriminant analysis model involves linear combinations of the following form:

$$Z = W_1X_1 + W_2X_2 + W_3X_3 + \dots + W_nX_n$$

where

Z = Discriminant score

W = Discriminant coefficient or weight for independent variable i

X = Predictor or independent variable i

The coefficients or weights (W) are estimated so that the groups differ as much as possible on the values of the discriminate function. This occurs when the ratio of between- group sum of squares to within- group sum of squares for the discriminant scores is at a maximum. Any other linear combination of the predictors will result in a smaller ratio.

In other words, discriminant score is to identify the roles of the variables loaded into the functions, where the bigger variables score will have a larger impact onto the loadings. Once the discriminant analysis model been computed, statistical test will be taken to evaluate the significance by using Wilk's (λ), or sometime known as U-statistics. The definition of several statistics associated with discriminant analysis as discussed by Malhotra N.K., 1999 (Malhotra N.K., 1999) will be reviewed in Appendix 3.

3.3.2.2 Factor Analysis

The next analysis is to identify the underlying reasons for the brand preference. This can be done by analyzing each brand's attributes, which is closely related to the reason of selecting a particular brand. These associations are consumer's behavior, and were recorded by using five points' scale in the questionnaire. Where, the value of one will be given if the respondent thinks that the attributes are highly associated with the brand, and five as least association.

The relevant attributes came from another qualitative research prior to the administration of this survey. As there are many attributes in the questionnaire, it is usually hard to administer or report all of them for each brand. Hence, factor analysis is chosen to help reduce the variables to a few manageable dimensions. This is done for each individual brand (Brand Usage) to understand the reason (Attributes Variable) for its preference. Apart from that, statistical test will also be carried out on each factor analysis processes to verify the accuracy.

Each variables in factor analysis is expressed as a linear combination of underlying factors. The covariation among the variables is described in terms of a small number of common factors plus a unique factor for each variables. The factors are not overtly observed. If the variables are standardized, the factor analysis model may represented as:

$$Z_1 = a_{11}F_1 + a_{12}F_2 + a_{13}F_3 + \ldots + a_{1m}F_m$$

$$Z_2 = a_{21}F_1 + a_{22}F_2 + a_{23}F_3 + \ldots + a_{2m}F_m$$

$$Z_j = a_{j1}F_1 + a_{j2}F_2 + a_{j3}F_3 + \ldots + a_{jm}F_m$$

where

Z_1 = the p-th standardized variable

a_1 = factor loadings

F_1 = common factor

m = number of common factors to all the variables

Eigenvalues will be use to identify the number of groups required in a particular factor analysis model, together with Kaiser- Meyer- Olkin (KMO) and Bartlett's test of sphericity are the modes of determining factor analysis measurement to the whole correlation matrix. The key statistics associated with factor analysis are as explained by Malhotra N.K. (1999) are indicated in Appendix 4.

3.4   Chapter Summary

The descriptive analysis, brand ranking, and the multivariate analysis findings will enable us:

1.   Understand market dynamics on the brand comparative position and ranking

2.   Identify differences between the prepaid and post paid user groups

3.   Further understand the customers' needs and wants, by the attributes for each brand

In short, Chapter 3 indicates the research methodology in relation to the survey objectives and literature review that were discussed in previous chapters. The summary of the workflow can be clearly identified in Figure 3.2.

Figure 3.2: Chapters Summary

Figure 3.2: Chapters Summary
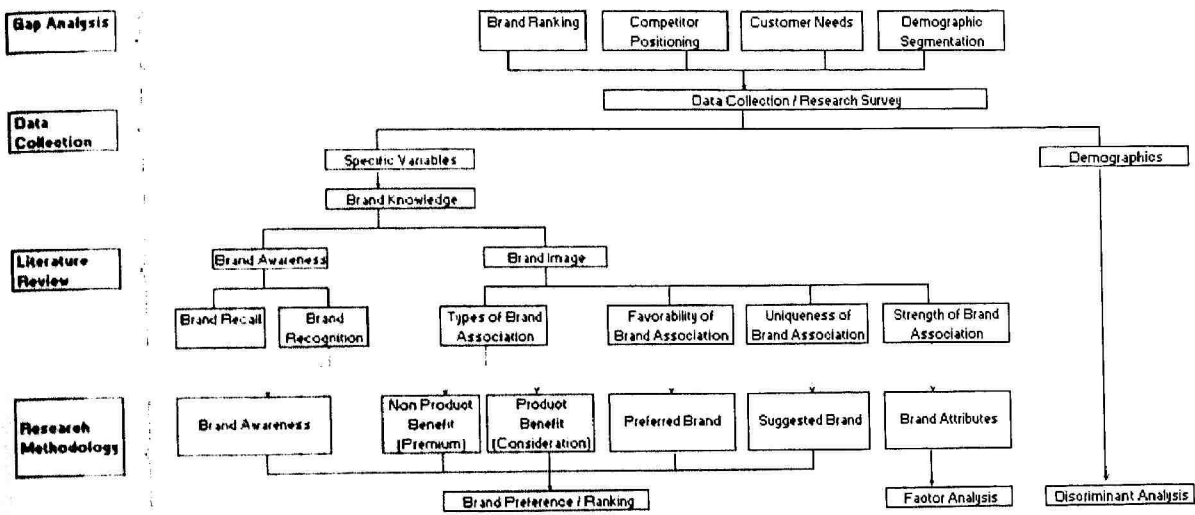
These results will help to answer most of the questions regarding mobile telecommunication units in the Malaysian market related to the objective of the survey. The next two chapters will report the detailed findings of the survey. It is separated into two chapters, as Chapter 4 reports on the general findings of survey, and Chapter 5 on more in- depth findings mostly on multivariate analysis and its tests.

CHAPTER 4: EXPLORATORY DATA ANALYSIS

4.1   **Demographic and Background Information**

4.2   **Brand Usage by Demographic Distribution**

4.3   **Brand Perception**

         4.3.1   **Brand Perception – Nokia**

         4.3.2   **Brand Perception – Motorola**

         4.3.3   **Brand Perception – Ericsson**

         4.3.4   **Brand Perception – Samsung**

         4.3.5   **Brand Perception – Alcatel**

4.4   **Comparison of Brand**

4.5   **Chapter Summary**