

Chapter 4: Research Methodology

4.1 Research Design

This study involves primary data collection through survey. The questionnaires were distributed and collected in the span of 3 weeks, where students were approached at certain strategic places such as the entry of Main Library and their respective faculty's libraries.

4.1.1 Sampling and Data Collection Procedure

This research adopted convenience (haphazard) sampling approach to get the representative of total student population. A questionnaire was administered to 302 students of various disciplines of study in University of Malaya, Kuala Lumpur. Each questionnaire was accompanied by a cover letter assuring anonymity and explaining the purpose of the study. The questionnaire was pre-tested on 15 students of various majors. Feedback was then used to refine the questionnaire. Peer review was also sought to further improve the questionnaire's clarity and readability. The respondent participating in the pre-test was excluded in the final sample.

4.1.2 Questionnaire Development

The questionnaires consisted of 3 parts (see Appendix C). In section 1, we asked respondent to indicate their demographic background (age, gender, marital status, religion, race, educational level, major of study), whether they own a computer and if not, where they have access to a computer most frequently, and finally self evaluation on their computer knowledge using a 5-point Likert type scale. Section 2 deals with the respondents' consumption experience, in which they were asked, among others, on their

willingness to pay for ODP, total ODP items purchased in the last 6 months, main reason for both buying and using pirated software, and whether they made illegal copy of software. Respondents were asked to indicate how many copies of pirated software they used/ currently use. Instead of asking them directly, they were instead required to select as many as appropriate from the list of softwares, with a designated space for them to write on any software that might not be on the list. Doing so gives clear advantages because when respondent are asked directly they are more likely to conceal the extent of their unethical behavior (Sims, et al 1996). Apart from that, we can also get the information on the extent of piracy for each type of the software.

4.2 Data Analysis Technique

The data was then analyzed using statistical tools such as frequencies and percentages, which were used in many sections. Cross tabulations were used to associate variables between sections and within a section. Several tests were also conducted. Among them are Mann-Whitney U Test, Wicoxon Signed Rank Test and Logistic Regression.

4.3 Econometric Specification

We are interested in examining the pirated ODPs buying decision of various categories of tertiary students as a function of several socioeconomic-demographic variables. The following conceptual model can summarize the theory behind our analysis:

Buy Pirated ODPs= f(demographics, computer ownership and computer knowledge, perception)

In our model, the propensity of *buying* pirated ODPs can be written as

$$Z_i=B_iX_i+\mu_i \qquad \dots\dots(1)$$

where

$$Z_i = \begin{cases} 1 & \text{if } \textit{buy and use} \text{ pirated ODPs} \\ 0 & \text{if otherwise (either } \textit{use but did not buy} \text{ or } \textit{did not buy or use}) \end{cases}$$
$$\text{Prob (buying pirated ODPs)} = \frac{e^z}{1 + e^z}$$

or equivalently

$$\text{Prob (buying pirated ODPs)} = \frac{1}{1 + e^{-z}}$$

where z is the linear combination

$$Z = b_0 + b_1 \textit{AGE} + b_2 \textit{GENDER} + b_3 \textit{MS} + b_4 \textit{REL1} + \dots + b_7 \textit{REL4} + b_8 \textit{ETHNIC1} + \dots + b_{10} \textit{ETHNIC3} + b_{11} \textit{EDCN} + b_{12} \textit{DISC1} + \dots + b_{14} \textit{DISC3} + b_{15} \textit{COMP} + b_{16} \textit{CKNOW} + b_{17} \textit{PRCP} + m$$

where;
AGE=Age of Respondents
GENDER=Gender of Respondents
MS=Marital Status
REL1=Buddhist
REL2=Hindu
REL3=Christian
REL4=Others
ETHNIC1=Chinese
ETHNIC2=Indian
ETHNIC3=Others
EDCN=Educational Level
DISC1=Science and Technology
DISC2=Business and Economics
DISC3=Arts and Social Sciences
COMP=Computer Ownership
CKNOW=Level of Computer Knowledge
PRCP=Perception and
m is the stochastic error term.

The variables used in the empirical analysis are described in Table A1 (Appendix A). Specification for the design variables for race, religion and discipline are described in Table A2 (also in Appendix A). Several variables have been recoded into new categories in order to avoid having too many independent variables and for the ease of analysis. Besides, total respondents for some of these variables are too small. These variables are faculty⁶, religion⁷ and race⁸. Summary statistics for the new categories were shown in later chapter.

⁶ Faculty was recoded by regrouping Education, Laws, Language and Linguistic, Art/Social Sciences, Sports Science, Islamic Studies into new discipline, DISC3 (Arts/Social Science). Economics and Administration, Business and Accountancy were combined to form DISC2 (Business & Economics). Science, Engineering, Build Environment, and Dentistry, Medical, & Nursing were grouped into DISC1 (Science & Technology). Computer Science retains its original form.

⁷ Religion has been recoded by grouping Sikhism, Taoism, Pagan, Free Thinker and Others into REL4 (Others).

⁸ Sarawak's Indigenous, Sabah's Indigenous, Orang Asli and Others were grouped into ETHNIC3 (Others)