# CHAPTER 3

## Methodology

### 3.1 Introduction

This study analyses the factors influencing the academic performances of upper secondary school students in the Sijil Pelajaran Examination (SPM) from four states. The survey includes responses of fifth and sixth formers from public secondary schools in Kuala Lumpur, Selangor, Kelantan, and Johor. This chapter describes the data and statistical method used to analyze the data. A general description of the students, the socio-economic characteristics of their parents, examination results, and other related variables that may affect academic achievement in specific subject areas are also discussed.

Data is analyzed using descriptive statistics, applying both graphical and numerical techniques. Chi-square test, analysis of variance, multiple and logistic regressions are used in the detailed analysis.

### 3.2 Data

The main source of data is obtained from a survey conducted in 1989 for the University of Malaya's tracer study on "Transition form School to Work" (See Chew et al., 1995 for details). This survey is the first part of a longitudinal study, which was conducted on two samples, consisting of Form 5 and Form 6 students.

33

In order to fulfill the objectives stated in Chapter 1, the study will focus only on the sixth formers, who would have both the SRP and SPM examination results. The performance in SRP examination indicates the prior achievement of students in each of three subjects at SPM examination. Relevant socio-demographic and socio-economic background of respondents, their academic achievement and their educational and career aspirations were captured in this survey in the form of open-ended and close-ended questions. (See Appendix 1 for the questionnaire used for data collection).

This research paper focuses on the data from the first part of a longitudinal survey following a cohort of secondary students from four states over a period of three months, from May 1989 to July 1989. The sampling frame of the study consists of secondary schools offering fifth and sixth form classes in four states in Malaysia, namely Kelantan, Johore, Federal Territory of Kuala Lumpur and Selangor. These states represent different level of economic growth. Kelantan represents the low-income state, Johore, the middle-income state, and the Federal Territory of Kuala Lumpur and Selangor represent the high-income states.

Prior to the survey proper, the researchers conducted a pilot survey on 27 students in Seremban and Kuala Lumpur. This pilot survey helped in identifying weaknesses in the original questionnaire. It was found that students were not willing to provide full details of their examination results, especially for those subjects where they had performed badly. This problem was overcome by instructing the survey administrators to inform the students of the importance of providing their detailed results.

The selected schools provided full support and cooperation during the course of this survey. The principals of the respective schools were also supportive by providing necessary information regarding their school. The survey yielded unweighted samples of 7944 fifth formers and 3283 sixth formers.

The sampling method used to obtain the sample was the disproportionate stratified sampling method. Stratification of the samples was done according to state, place of stay (urban-rural), in stream of study (academic, technical, vocational and residential stream) and school size. As a disproportionate stratified sample was used, assigning of appropriate weightings to the sample was necessary. The weighted samples consist of 8141 and 2867 fifth and sixth form students respectively. The detailed descriptions related to the survey can be found in Chew et al. (1995).

## 3.3    Definition of Terms and Variables

### 3.3.1    Academic achievement

Academic achievement is measured by the grades obtained in the National public examinations known as the Sijil Pelajaran Malaysia (SPM). Secondary school students take this examination at the end of their fifth year of secondary education. The subjects taken by students at this examination vary according to their stream of study but there are some core subjects, which are compulsory for all students such as Bahasa Malaysia, English and Mathematics.

This report is concerned with the achievement of students in three core subjects, namely, Bahasa Malaysia, English and Modern Mathematics. Subject performance in SPM examination is graded on an alphanumeric scale from A1 (at the top of scale) to F9. This has been converted to a numerical scale from 1 to 9. The grades for each subject in the SPM examination is then further classified into Distinction (A1, A2), Credit (C3, C4, and C5), Pass (P7, P8), or Fail (F9)

### 3.3.2 Stream of Study

**Table 3.1: Distribution of Students by Stream of Study and Gender**

| | Stream of study (%) | | | | Total |
|---|---|---|---|---|---|
| | Science | Arts | Art-Commerce | Technical | |
| Male | 35.3 (427) | 62.9 (761) | 1.8 (22) | - | 100(1210) |
| Female | 15.3 (254) | 82.2 (1362) | 2.2 (37) | 0.2 (4) | 100(1657) |
| Total | 23.8 (681) | 74(2123) | 2.1 (59) | 0.1 (4) | 100(2867) |

Note: Number of students for each cell is in parentheses. (-) Indicates a percentage less than 0.01

Students included in the survey represent four streams of studies, with vast majority from the Arts/ Art-Commerce stream (See Table 3.1). About less than a quarter are science stream students. For the purpose of this study the Art-Commerce students, who are of a small sample size will be merged, with the Arts students. Since three out four students in the technical stream have done non- science subjects, they too will be combined with the Arts group and therefore, students are classified as Science or Arts only in this report. The total number of students representing these two streams is 681 and 2186 respectively.

### 3.3.3 Location of Primary School and Childhood Residence

In the survey the place where students grew up in is classified as rural, small town and large town. The term "rural" refers to those schools located in what is generally referred to as rural and remote areas of the four states name Kuala Lumpur, Selangor and Kelantan. These schools could be located in villages or new land development scheme. Small towns may not have the facilities of large towns but are generally more developed than rural areas. Similarly, the place where these students grew up is classified in the same manner. Students growing up, studying and living in rural areas have many different social, cultural, economic and educational pursuits.

### 3.3.4 Socio-Economic Status (SES)

As supported in several studies (Theisen, 1983; Kutnick and Jules,1988), SES, which is a measure of parental income, occupation and level of education was found to be a significant predictor of academic achievement in developing countries. In this study, the SES is reflected by father's educational attainment and income for respondents whose guardian is either both parents or a single father. For respondents whose guardian are a single mother, the educational attainment and income of the mother is taken into consideration while the educational attainment of the guardian is considered if the respondents are staying with a guardian.

### 3.3.5 Dependent and Independent Variables

The dependent variable is the examination grade (A1... F9) obtained for each of the three subjects at the SPM examination. The level of achievement (high) for each of the three selected subjects will be used as the dependent variable in logistic regression. The socio-demographic, socio-economic and educational background of the students' and its interactions are the independent variables of interest and these are divided in three groups as listed in Table 3.2

**Table 3.2: Selected Independent Variables Used in the Analysis**

| 1.Demographic Variables | |
|---|---|
| Gender | Female<br>Male |
| Ethnicity | Bumiputra[1]<br>Chinese<br>Indians |
| Childhood Residence | Large Towns<br>Small Towns<br>Rural |
| Birth Order | Eldest<br>Second / Third<br>Fourth / Fifth<br>Sixth or Higher |
| **2.Home Environment** | |
| Family Size | Small (1-2 children)<br>Average (3-4 children)<br>Large (5 or more) |
| Education level of parents' | No formal education<br>Lower secondary<br>Upper secondary<br>Tertiary (College/ University) |
| Income of parents' | Less than RM1000<br>RM1001-RM1500<br>RM1501-RM2000<br>RM2001 or more |
| Occupation sector of parents' | Agricultural<br>Private sector[2]<br>Civil Service |
| Importance of doing well to parents | Important<br>Not important |
| How often parents check school work | Often<br>Occasionally |

| 3.Educational Information | |
| --- | --- |
| Location Primary school attended | Large Towns<br>Small Towns<br>Rural |
| [2]Medium of instruction in primary school | Bahasa Malaysia<br>Chinese |
| Examination results of SRP and SPM | Distinction (A1, A2)<br>Credit (C3, C4, C5, C6)<br>Pass (P7, P8)<br>Fail (F9) |
| Additional tuition taken | Yes<br>No |
| Stream Of Study | Science<br>Arts |
| Frequency of library usage | Often<br>Rarely |
| Frequency of reading Newspaper/Books | Daily<br>Occasionally |
| Commonly used language for reading | Bahasa Malaysia<br>English<br>Chinese |

Note: [1] Includes Malays and other Bumiputras'. [2]Students from Tamil, English and other medium of instructions are excluded due small sample size (<1 percent). [3] Combines mining, construction, electricity, manufacturing, electricity, commerce, transport due to small sample size of each sector.

## 3.4 Statistical Techniques

(i) Exploratory Data Analysis

The exploratory data analysis consists of graphical displays such as pie charts and bar charts as well as summary of statistics for each variable. The graphical display enables us to see the data as a whole and examine the patters of the distribution. Numerical summaries (refer to Berenson, Seventh Edition) included in the sample are mean, standard deviation coefficient of variation, skewness and the size of the sample. Cross tabulations provide an overview of the differences in the grade distribution of the three subjects among students from various backgrounds.

(ii) Chi-Square Analysis

It is also of interest to study association between two categorical variables. In these situations, Chi-Square test of independence is used as a confirmatory analysis (refer to Keller Warrack). In this study, Chi Square analysis will be used to see if there is significant association between respective subject grades and selected independent variables. The null and the alternative hypothese would be:

$H_0$: The two categorical variables are independent (i.e. no association between variables)

$H_1$: The two categorical variables are not independent (i.e. there is an association between them).

41

The decision is to reject the null hypotheses if the computed Chi Square statistics is greater than the critical value from the Chi-Square distribution with $(r-1)(c-1)$ degrees of freedom, where r =row total and c= column total or if the p-value of the Chi square statistics is less than the predetermined level of significance which is 0.05 in this study. It would be necessary to combined categories if cells have frequencies below 5 for Chi-Square tests to be valid.

(iii)    Multiple Regression

Regression analysis is a statistical technique that is used to predict the values of a dependent variable based on values of at least one explanatory or independent variables. In this study, the Multiple regression models in which several explanatory variables are used to explain the changes in the dependent variables will be considered. Multiple regression is used to identify the factors that will influence the performance of students in each of the three subjects. In other words, three regression equations with a set of independent variables for each subject will be obtained. The grades obtained in English, Bahasa Malaysia and Mathematics for the SPM examinations is used as the dependent variable respectively.

The general multiple regression model is as follows:

$$Y = \beta_0 + \beta_1 X_1 + ... + \beta_k X_k + \alpha_1 D_1 + ... + \alpha_n D_n$$

where Y is the dependent variable (estimated grade for a particular subject) and $X_1...X_k$ represent independent variables with numeric values (subject grades and family size), while $D_1...D_n$ represents categorical independent variable (ethnicity, childhood residence, stream)

(iv)    Logistic Regression

Logistic Regression analysis is an extension of multiple regression analysis techniques used to research situations in which the dependent variable is dichotomous (outcomes with two categories). In analyzing student performance in the three subjects, logistic regression allows the predictions to be made of the outcome of a student being a high achiever or non- high achiever. Students who scored A1, A2 and C3 for each of the three subjects are classified as high achieving students in this study.

The model for logistic regression used in this study is:

$$E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \ldots\ldots + \beta_k X_k \text{ where,}$$

Y is the outcome variable taking the value of 1 (i.e. the positive outcome, high achievers) and 0 (i.e., the negative outcome, non high achievers) with a set of k predictor variables that are significantly related to Y which, provides additional information for predicting Y. The Logistic regression analysis is based on a linear model for the natural logarithm of the odds (i.e., the log-odds) in favour of Y = 1.