CHAPTER 2

RESEARCH METHODOLOGY

2.1 DATA SOURCES

The data for this research came from the Malaysian Population and Family Survey (MPFS) conducted by the National Population and Family Development Board (NPFDB) in 1994/95. The objectives of the 1994/95 MPFS were to provide essential information on marriage, childbearing, family planning and some other aspects of family life to the relevant agencies in formulating and implementing an integrated population and family development programs.

A stratified random sample, with probability proportional to size of population, was drawn for the survey. The sample was stratified according to urban-rural location, based on the updated sampling frame maintained at the Department of Statistics (Malaysia). The main survey was conducted by the NPFDB from September 1994 to February 1995. A total of 6,015 households in Peninsular Malaysia were selected for the survey. Upon screening, a total of 4,444 ever married women aged 15-49 years were selected as eligible respondents for detailed interviews. Of these, 4,240 were currently married women, 102 were widows, 74 divorcees and 28 women were separated from their husbands. In this study, childless women are excluded from the analysis on age at first birth. Out of 4,134 respondents who had ever given birth, 3,832 of their first births were live births, 43 stillbirths and 259 abortions.

Basic socio-demographic data from each household member was recorded in a household schedule. A host of open-ended and close-ended questions were used to collect detailed information from the ever-married women aged 15-49 years. Information collected from the survey include respondent's socio-demographic background, marriage history, pregnancy history, family planning knowledge and practices, work history, husband's background and the care of the elderly. While it is common practice to examine women's characteristics in socioeconomic study of fertility, the views and behavior of their spouses are also important in the timing of marriage and number of children they have. For that reason, the 1994/95 MPFS collected some information about the husbands.

Comparing the sample distribution of the 1994/95 MPFS with that of the 1991 population census on a number of socio-demographic characteristics, it can be seen that the survey is rather representative of the national population. In terms of ethnicity, the proportion of Malay women in the sample corresponds very closely that of the general population. However, Chinese women were slightly under-represented while Indian women were over-represented (see Table 2.1). Part of this deviation from the general population may be due to the differences in marital structure, as the survey was confined to married women only, and that the proportion who had been married among Indian women was probably higher among that of Chinese women in the 15-49 age group.

Table 2.1: Sample and Population Distributions by Ethnicity

Ethnic Group	1994/95 MPFS (per cent)	Women Aged 15-49 from the 1991 Population Census (per cent)
All races	100.0 (4444)	100.0
Malays	59.7	59.4
Chinese	24.8	29.1
Indians	11.7	8.5
Others	3.8	3.0

Table 2.2: Sample and Population Distributions by Education Level

Education Level	1994/95 MPFS (per cent)	Women Aged 15-49 from the 1991 Population Census (per cent)
All Levels	100.0 (4444)	100
No formal education	8.0	12.1
Primary	37.8	28.7
Secondary	46.4	49.7
Tertiary	7.7	9.5

Comparing the 1994/95 MPFS data and 1991 Census data on education level of women, it appears that those with no formal education were slightly under represented and those with primary education appear to be over represented. However, the sample proportions that had secondary and tertiary education correspond rather closely with the 1991 Population Census (see Table 2.2).

Table 2.3 shows that the distribution of the sample according to place of residence corresponds very closely that of the general population, as reported in the 1991 census.

Table 2.3: Sample and Population Distribution According to Place of Residence

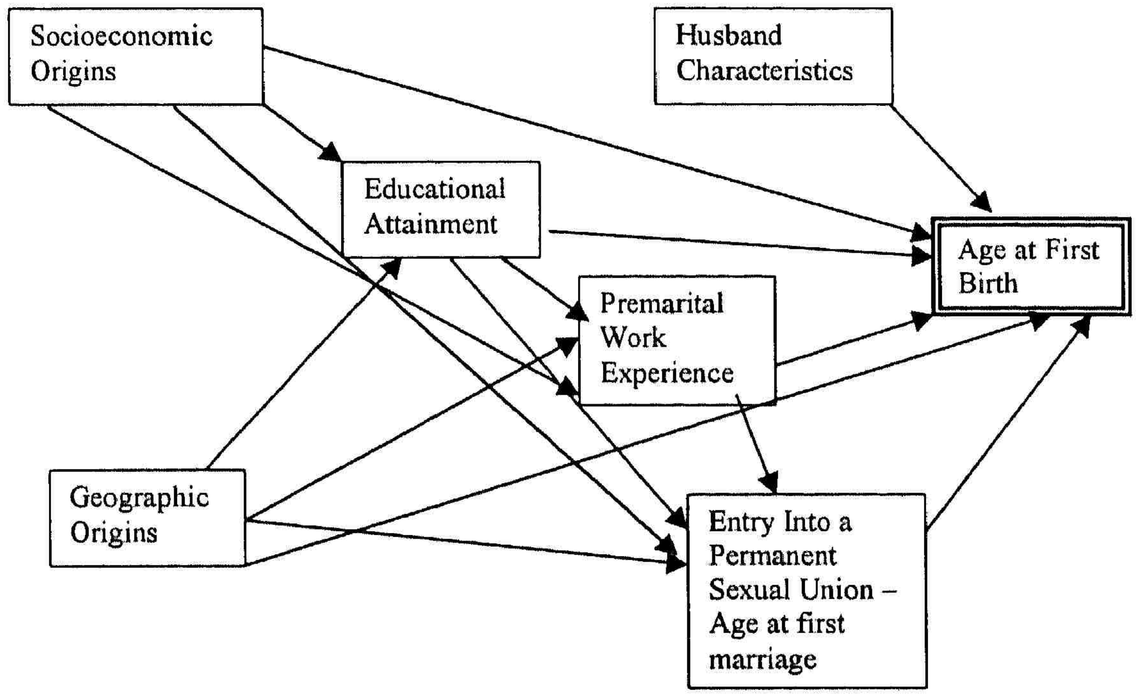
Place of Residence	1994/95 MPFS (per cent)	Women Aged 15-49 from the 1991 Population Census (per cent)
Total	100.0 (4444)	100.0
Rural	45.7	47.5
Urban	54.3	52.5

2.2 ANALYTICAL FRAMEWORK

This study is aimed at examining the social and economic determinants of age at first birth among married women in Peninsular Malaysia. The beginning of childbearing, as measured by the age of woman at the birth of her first child (defined as the first live birth) is the dependent variable for the analyses. It is obtained by subtracting the year and month of birth of the respondent in the survey from the year and month of her first live birth. In this survey, 14.8 per cent of the respondents reported year, but not month for one of these events. For all these cases, a mid-year date was imputed.

The analytical framework for the analysis of age at first birth is adapted from a causal model proposed by Hirschman, as shown in Figure 2.1.

Figure 2.1: Analytical Framework for Age at First Birth



The sequence of events in the life cycle has been used to order the events in the causal model. From the model, intervening between social and geographical origins and the outcome variables are indicators of women's participation in modern society – education and early work experience (Hirschman and Rindfuss, 1980). With higher education, women are more likely to break away from traditional values and behaviors, which include early marriage and childbearing. Modern women have greater say in the timing of family formation. In the following sub-sections, the relationship between age at first birth and some selected variables, as documented in previous studies will be examined.

The 1994/95 MPFS provides a rich source of data for examining the changing trends and patterns of age at first birth across birth cohorts in terms of a number of socio-economic variables. In this data set, age at first birth is approximately normally distributed, and hence the mean and median values are about the same. This research will examine the variations in age at first birth among cohorts of women with similar background and socioeconomic characteristics, where the same social structural norms apply (Jones, 1994:62).

To minimize the problems of censoring and truncation, the sample is restricted to ever-married women above 30 years with one or more live births (Rindfuss and others, 1983, pp.258-260). This reduces the sample size to 2748, with the Malays making up 58.4 per cent of the sample, Chinese 28.9 per cent, Indians 10.0 per cent and others the remaining 2.7 per cent.

The selection of independent variables is guided by previous studies on the determinants on age at first birth. Past studies showed wide variations in age at first birth across ethnic groups. In this study, variation in age at first birth among Malaysia's three main ethnic groups (the Malays, the Chinese and the Indians) will be examined. Age at birth has been found to vary with childhood and current place of residence. Urbanization is posited to be an important influence on the timing of first birth because of the broader range of ideas and roles that are open to young women in urban than in rural environment. Other factors including the education levels and work status of both the spouses also provide sharp differentials in age at first birth and will be examined in this study.

Husband's educational level and spousal age differences will be examined to provide information on their effects on the wife's age at first birth. Frequency distribution of the independent variables is shown in Table 2.4.

Table 2.4: Sample Distribution of Ever-Married Women Aged 30 Years and Above with at Least One Live Birth According to Selected Independent Variables

Variable	Distribution		Variable	Distribution	
	n	%		n	%
Total	2748	100			
*Ethnicity**			*Ever worked before*		
Malay	1604	58.4	*marriage*	847	30.8
Chinese	794	28.9	No	1900	69.2
Indian	276	10.0	Yes		
Place of Residence			*Husband's Education*		
Urban	1481	53.9	No formal education	126	4.6
Rural	1267	46.1	Primary	1178	42.9
			Secondary	1175	42.8
			Tertiary	269	9.8
Birth Cohort			*Age difference between*		
Before 1951	715	26.0	*spouses*		
1951 - 1955	591	21.5	Wife older than husband	461	16.8
1956 - 1960	804	29.3	0 to 5 years	1314	47.8
1961 - 1965	638	23.2	6 to 10 years	733	26.7
			More than 10 years	240	8.7
Wife 's Education					
No formal education	267	9.7			
Primary	1251	45.5			
Secondary	1048	38.1			
Tertiary	182	6.6			

* Exclude Others

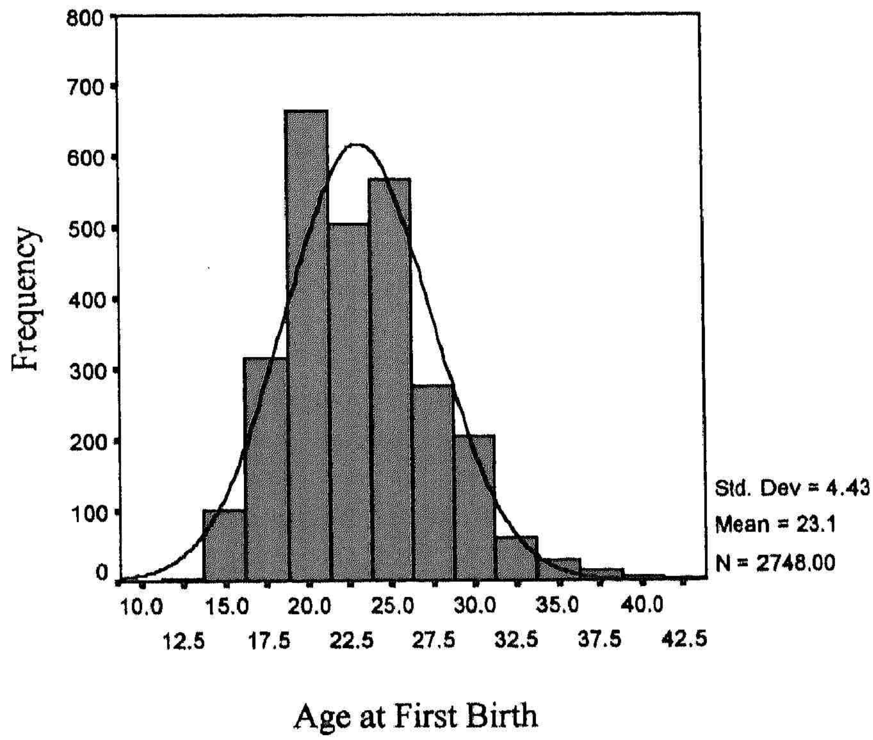
2.3 STATISTICAL TECHNIQUES USED IN THE ANALYSES

The 1994/95 MPFS shows that the mean age at first birth among Malaysian women is about 23 years. Half of the women had their first birth before 23 years and the remaining half after age 23. Data show that age at first birth is approximately normally distributed with skewness of 0.599 (see Table 2.5 and Figure 2.2).

Table 2.5: Summary Statistics of the Age at First Birth Among Ever-Married Women Aged 30 Years and Above with at Least One Live Birth

Age at First Birth		
N	Valid	2748
Mean		23.06
Median		23.00
Std. Deviation		4.43
Skewness		.599
Std. Error of Skewness		.047
Kurtosis		.451
Std. Error of Kurtosis		.093

Figure 2.2: Histogram of Age at First Birth Among Ever-Married Women Aged 30 Years and Above with at Least One Live Birth



Bivariate analyses will be performed to examine the relationship between age at first birth and selected socioeconomic variables like ethnicity, place of residence, woman's educational level, premarital work experience, husband's educational level, age difference between spouses and age at first marriage.

Several multivariate techniques will be used to disentangle the complex relationships among the independent variables and age at first birth. As age at first birth is very strongly correlated with age at first marriage, the effects of socio-demographic factors on age at first birth will be examined with and without the inclusion of age at first marriage as an intervening variable. This will allow us to examine the extent to which the effects of these variables are mediated through age at marriage. Multiple regression and multiple classification analysis will be used to build the multivariate models to determine the factors affecting age at first birth. ANOVA will be performed to evaluate if there are significant differences in the mean age at first birth across groups.

Multiple Regression

Multiple regression is a statistical tool used to analyze the linear relationship between age at first birth and a set of independent/predictor variables. As an analytical tool, it is used specifically in this study to examine the effects of each independent variable on age at first birth, controlling for other confounding variables,

Multiple Classification Analysis

Multiple classification analysis (MCA) is a technique for determining the interrelationships between a dependent variable and several explanatory variables within the context of an additive model. Essentially, it is a variant of multiple regression technique using dummy or binary variables. Its main advantage over the dummy variable regression technique is that outputs are easier to interpret, because the coefficients expressed as deviations from the grand mean, instead of deviations from a constant term based on a composite sum of means of the excluded subclasses. Another advantage of MCA is nominal-scale predictor variables can be entered without recoding. It also provides useful summary measures such as beta for each predictor to indicate it's relative importance after adjusting for effects of other predictors, and R^2 which is the proportion of the total variance in the dependent variable accounted for by all the predictors in the model.

Logistic Regression

Logistic regression is used to predict the presence or absence of a characteristic or outcome based on values of a set of predictor variables. It is similar to a linear regression model but is suited for models in which the dependent variable is dichotomous. Logistic regression coefficients can be used to estimate odds ratios for each of the independent variables in the model. In this analysis, the value that is being predicted represents a probability, and it varies between 0 and 1. The parameters of the model are estimated using the maximum-likelihood method. That is, the coefficients that make our observed results most 'likely' are selected.

Multiple regression and multiple classification analyses are used to construct models that relate age at first birth to all the independent variables. The explanatory variables tested are the woman's own characteristics, age at first marriage and husband's characteristics. All the background variables, except age at first marriage, are entered as sets of dummy variables. For all the dichotomized dummy variables used, the variable assigned a value of zero will be used as the reference category. In the case of ethnicity, Indians is used as the reference category. Place of residence has been recoded into dichotomous variable with '1' for urban areas and '0' for rural areas. Pre-marital work is dichotomized into two categories namely code '1' for those who worked on a paid job, and '0' for those who never worked before or worked as unpaid family worker.

Education, as measured by years of schooling, is a powerful predictor of the timing of first birth. Husband and wife's educational levels are entered as categorical variables in the multiple classification analyses. The level of education is classified as follows:

- No formal education
- Primary - referring to those with 1-6 years of education
- Secondary - referring to those with 7-11 years of education
- Tertiary - referring to those with more than 12 years of education

In the regression analyses, education levels are entered as dummy variables, with tertiary education as the reference category. Age differences between spouses are measured by taking the differences between the wife's ages from the husband's age. If the wife is older than the husband, then the age difference will be negative. In the regression analysis, the age difference is entered in the model in ratio scale, in terms of number of years. In multiple classification analyses, the age differences are divided into four categories namely, negative values (wife older than husband); 0 to 5 years, 6 to 10 years and more than 10 years.

In this study, logistic regression is conducted to analyze the relationship between the probability that a woman would give birth by age 25 years and the independent variables listed earlier.