

CHAPTER 2

DATA AND METHODOLOGY

2.1 Introduction

This research is based on data from a survey conducted in 1997/98 by a group of researchers from Universiti Kebangsaan Malaysia on behalf of Malaysia Tourism Promotion Board (MTPB). The survey collected detailed information pertaining to the characteristics of domestic tourists, their travel patterns and the favourite tourist destinations, to provide the necessary inputs for the formulation and implementation of marketing strategies to enhance domestic tourism in Malaysia.

This chapter describes the research methodology in terms of conceptual and operational definitions, data collection method and sampling design. The analytical framework and the main study variables will also be described.

2.2 The survey methodology

2.2.1 Definitions of the key concepts

Researchers may adopt different terminologies and operational definitions of the study variables. At this juncture, it is necessary to describe the conceptual and operational definitions of the main study variables used in the survey.

A **trip** refers to a journey made at least 40 kilometres (one way) from respondent's place of residence regardless of whether it involves an overnight stay. However, it does not include daily journey to the work place.

A **domestic tourist** is a person residing in Malaysia, irrespective of nationality, who travels to a place at least 40 kilometres (one way) from his/her usual place of residence for leisure or holidays.

Destination means the place where the respondent spends the longest time and the place furthest from the respondent's home for any purposes, e.g. vacation, business, visiting relatives or friends, etc.

Non-package tour is self-arranged trip, while **full-board package** means the trip is fully organized by the organizer/employer/travel agent. A **half-board package** is a trip partially organized by the organizer/employer/travel agent.

2.2.2 Sampling Designs

The study population for this survey comprises residents of Malaysia living in 14 major towns for at least 36 months prior to August 1998; with a monthly household income of RM750 or more. A stratified random sample was selected for the survey. The households in the major towns were stratified according to region and ethnicity. This gives a total

of 42 strata (14 towns x 3 ethnic groups). Within each stratum, a simple random sample was selected.

The sampling frame for the survey was based on the listings of the Department of Statistics for the 1991 population census. Table 2.1 shows the distribution of the total number of households (as of 1991) and selected households in selected towns. Of the total 6,971 households selected for the survey, only 4,236 returned the completed form at the end of the survey period, yielding a response rate of 60.8 percent.

Table 2.1: Distribution of total number of households (as of 1991) and selected households in selected towns

Selected Towns	Number of Households	Number of Selected Households	Number of Households Completing the Survey
1. Johor Bahru	105,016	682	339
2. Malacca	33,171	215	148
3. Seremban	33,878	224	224
4. Klang Valley (with KL)	469,791	3053	1,624
5. Ipoh	95,795	623	306
6. Alor Setar	56,558	367	343
7. Penang	76,629	498	155
8. Kangar	8,571	110	110
9. Kota Bharu	48,780	329	329
10. Kuala Terengganu	34,414	224	183
11. Kuantan	33,896	220	158
12. Kuching	39,912	259	256
13. Kota Kinabalu	19,972	130	24
14. Labuan	4,505	37	37
Total	1,060,888	6971	4,236

The low response rate is due to failure on the part of some respondents to cooperate. Burns and Busch (1997) and Scheaffer, Mendenhall and Ott (1996) have noted the low response rate with the diary panel method and self-administered questionnaire. The research assistants, who were selected from each town, also played an important role in ensuring the success of the survey. They were briefed on the procedures and criteria of selecting respondents and methods of collecting information from the respondents. The research assistants were instructed to revisit the respondents and complete missing information.

2.2.3 Methods of data collection

A questionnaire was designed to collect detailed information regarding travelling of each household member. The questionnaire was modified based on a pre-test conducted in the Klang Valley area.

This study uses a diary panel method that recorded travel information of each household member for a 12-month period. Each selected household was given a diary (Appendix I) to record all information regarding visits made by household members between August 1997 and July 1998. This information was collected at the end of every month. Data received monthly were then edited and processed for further analysis.

A detailed manual was provided to assist respondents in filling the questionnaires. This manual also provides the operational definitions of the terms used in the study (Appendix II).

2.3 The Main Study Variables

The dependent variables for this study are:

1. Whether respondents took a domestic tour (coded as 1 if yes and 0 if not);
2. Frequency of domestic tours;
3. Types of tour (group, individual, family);
4. Destination;
5. Type of destinations (highlands/hill, beach/island, golf resort, theme parks, cities/town, etc); and
6. Expenditure

The independent variables for this study are:

1. Ethnicity;
2. Gender;
3. Age of respondent;
4. Marital status;
5. Educational level;
6. Individual income;
7. Household income; and
8. Region of origin.

Data show that out of 12,476 persons (aged twenty and above) from the selected households, only 4,687 had gone on a domestic tour. This shows that 62.4 percent had not travelled for vacation during the study period.

Very few in the study population had travelled more than twice.

Table 2.2: Distribution of the sample population by number of domestic tours during the year

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	7789	62.4	62.4	62.4
	1	2931	23.5	23.5	85.9
	2	1135	9.1	9.1	95.0
	3	398	3.2	3.2	98.2
	4	150	1.2	1.2	99.4
	5	57	0.5	0.5	99.9
	6	9	0.1	0.1	99.9
	7	3	0.0	0.0	100.0
	8	4	0.0	0.0	100.0
Total		12476	100.0	100.0	

Table 2.3 shows the frequency and percentage distributions of all the respondents and tourists according to selected independent variables. Besides showing the sample split of the study population, a comparison of the two distributions indicates the differentials in the propensity to travel for the various sub-groups of the population. Data show that there are about the same number of males and females in the sample and their propensity to travel is not significantly different. The Malays make up 57.6% of the sample but 61.1% of the tourists, indicating that they have higher propensity to travel as compared to that of other ethnic groups. The modal class for the respondents and travellers are those aged 35-44 years, and 75.9% were married. In terms of educational level, 51.7% of the respondents had secondary education, and 23.4% had tertiary education. The relatively higher percentage of tourists in these two educational categories as compared to their distribution in the general population indicates that the propensity to travel increases with

education, and this may be explained by their higher income. Nearly one quarter of the respondents were from households with a monthly income of RM750-1000 and about half were from households that had an income of RM1000-3000 a month. Data show that while 52.1% of the respondents had come from the central region, it accounted for only 42.5% of the tourists. This indicates that residents from the central region were relatively less likely to travel to other parts of the country as compared to those from other regions.

2.4 Statistical Analyses

The statistical techniques to be used in the data analysis are to a large extent determined by the scale of measurement and data distribution. In this report, the main dependent variables are cross-tabulated with socio-demographic variables to examine the patterns of domestic tourism, in terms of differentials by age, gender, ethnicity, educational level, marital status, income, region, etc.

As many variables are inter-related in a complex manner in explaining the propensity to travel, there is a need to examine the study variables within the multivariate context. Multivariate techniques such as logistic regression will be used to study the independent and combined effects of socio-demographic variables on domestic tourism. Appropriate tests will also be conducted to assess the statistical significance of the relationships of the study variables.

Table 2.3: Frequency and percentage distributions of the independent variables in the analysis

	Total Respondents		Total Tourists	
	Frequency	Percent	Frequency	Percent
Gender				
Male	6095	48.9	2307	49.2
Female	6381	51.1	2380	50.8
Total	12476	100.0	4687	100.0
Ethnicity				
Malay	7189	57.6	2866	61.1
Chinese	3195	25.6	1125	24.0
Indian	1490	11.9	516	11.0
Others	602	4.8	180	3.8
Total	12476	100.0	4687	100.0
Age				
20-24	2330	18.7	851	18.2
25-34	2113	16.9	816	17.4
35-44	4122	33.0	1634	34.9
45-54	3053	24.5	1107	23.6
>54	858	6.9	279	6.0
Total	12476	100.0	4687	100.0
Marital Status				
Married	9468	75.9	3608	77.0
Single	2834	22.7	1021	21.8
Others	174	1.4	58	1.2
Total	12476	100.0	4687	100.0
Educational Level*				
No Formal Education	686	5.8	203	4.5
Primary Education	2262	19.1	725	16.2
Secondary Education	6134	51.7	2350	52.4
Tertiary Education	2776	23.4	1204	26.9
Total	11858	100.0	4482	100.0
Individual Income*				
No Income	3917	32.5	1377	30.3
RM1-1,000	5328	44.2	1905	41.9
RM1,001-2000	2037	16.9	877	19.3
RM2,001-3000	469	3.9	228	5.0
More than RM4,000	314	2.6	155	3.4
Total	12065	100.0	4542	100.0
Household Income*				
RM750 -1,000	3025	24.5	1020	22.0
RM1,001-2000	3946	32.0	1415	30.5
RM2,001-3000	2237	18.1	859	18.5
RM3,001-4000	1362	11.0	599	12.9
More than RM4,000	1761	14.3	741	16.0
Total	12331	100.0	4634	100.0
Region of Origin				
Central	2444	52.1	5307	42.5
North	764	16.3	2300	18.4
South	507	10.8	1798	14.4
East	540	11.5	1701	13.6
Sabah & Sarawak	432	9.2	1370	11.0
Total	4687	100.0	12476	100.0

* Excluding not disclosed cases

The effect of each variable on domestic tourism is assessed net of the effects of other variables in the models. Diagnostic analyses will be done on the main study variables to check for serious violations of the normality assumption. Table 2.4 and Figure 2.1 and 2.2 show that the distribution of all respondents by number of vacations is not normally distributed, with those that have never travelled making up 62.4% of the total.

Table 2.4 shows that the number of vacations for all respondents and tourists is positively skewed (with skewness slightly in excess of 2 and kurtosis in excess of 5). The Kolmogorov-Smirnov tests show violation of the normality assumption. Therefore, parametric multivariate techniques such as multiple classification analysis, analysis of variance and multiple regression are inappropriate.

Table 2.4: Summary statistics of the number of vacations for all respondents

	Total Respondents	Total Tourists
Sample size, n	12476	4687
Mean	0.592	1.576
Skewness	2.061	2.037
Std. Error of Skewness	0.022	0.036
Kurtosis	5.452	5.324
Std. Error of Kurtosis	0.044	0.072
Kolmogorov-Smirnov Z	39.989	24.635
Asymp. Sig. (2-tailed)	0.000	0.000

Figure 2.1: Histogram and normal curve on the number of vacation for all respondents

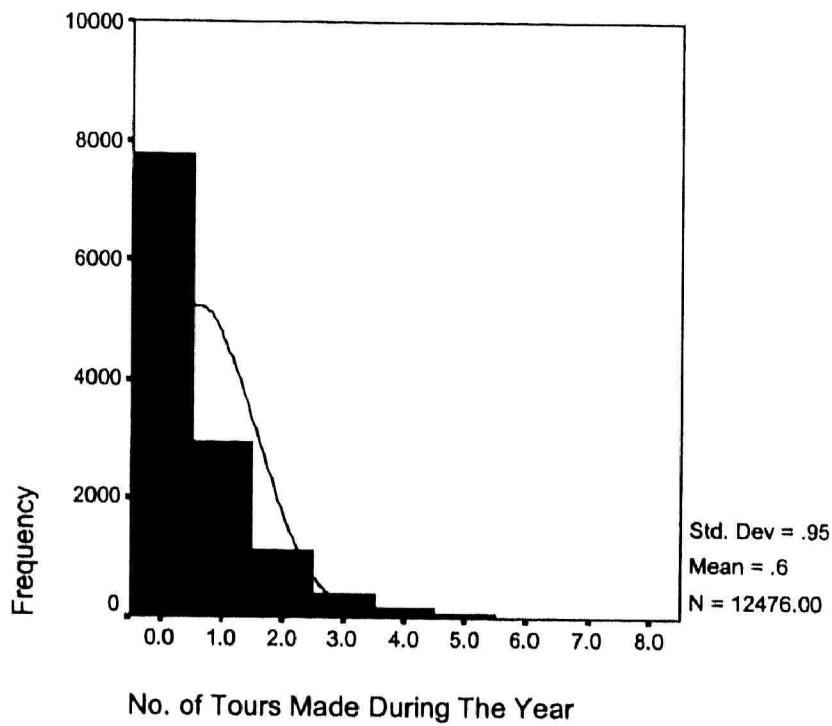
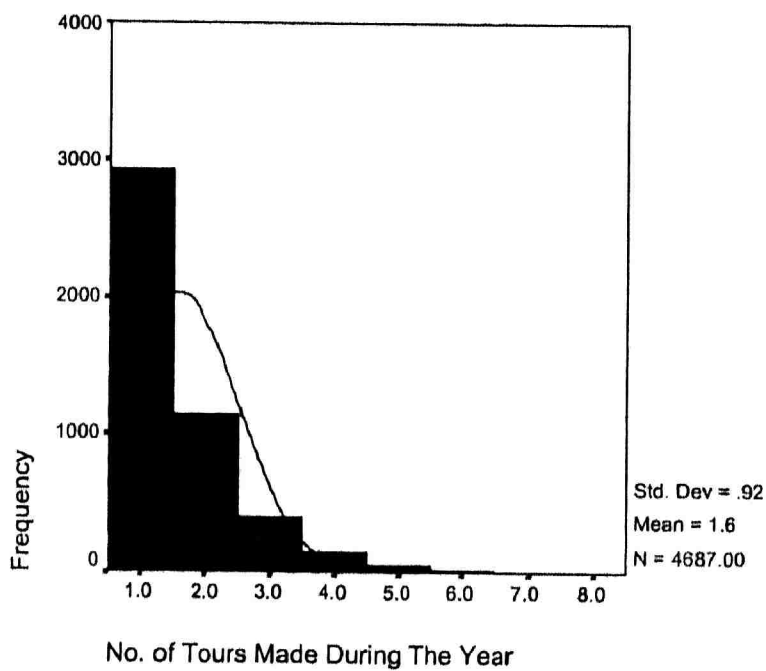


Figure 2.2: Histogram and normal curve on the number of vacation for tourists



Bivariate analyses are performed to examine the relationship between the number of vacations (for all respondents and tourists separately) and selected characteristics. Analysis of variance is performed to evaluate if there are significant differences in the mean number of vacations across the various sub-groups.

Logistic regression analyses are used to estimate models that relate the probability of domestic tourism with selected socio-demographic variables, such as gender, ethnicity, educational levels, marital status, household income and region of origin.