CHAPTER III

METHODOLOGY AND SOURCE OF DATA

Data Source and Objective of the Media Index Survey

In order to analyze the design effects of a public survey in Peninsular laysia, the sampling design must conform to the theoretical requirements of bability sampling methods. Probability sampling can provide valid estimates of unknown population values and allow for evaluation of the reliability of the ulting population estimates. Besides, the survey design should also be the fairly ndard and commonly used in the country to allow for generalization of the results future surveys.

The data used in this study is from Media Index Survey (MIS) which is a ntinuous survey fielded by ACNielsen Malaysia to monitor the media nsumption in Peninsular Malaysia. The survey incorporated a stratified multistage npling design. The overall sample is divided into 4 quarters of fieldwork periods th each quarter representing a random sample by itself. A nationally resentative sample is achieved quarterly with target sample size of 2200 based on :e to face personal interviewing. The survey was conducted between July 1997 d June 1998.

The main objective of the Media Index Survey is to obtain information on e media consumption habits among adults aged 15 years old and above in ninsular Malaysia to provide the necessary information as media currency to the edia industry. The survey elicited information on consumer behavior with respect printed materials, cinema, radio, television and video. Besides, the survey also

ected a host of information on household and personal product usage as well as
iographic information.

Sampling Design of Media Index Survey (MIS)

The sampling design adopted for Media Index Survey is a stratified 2-stage
iter sampling with the target population of adult aged 15 years old and above.
: survey covered urban and rural areas of all the 11 states in Peninsular Malaysia
the Federal Territory of Kuala Lumpur. Sabah and Sarawak were not covered in
survey for reasons of high costs of the survey, and lack of demand for these
irmation from the market in these areas. A total of 8738 respondents were
irviewed from 439 PSUs across the peninsula.

The survey was conducted utilizing area sampling frame, which is the widely
d sampling frame especially in general public surveys, where the ultimate
ipling element is person in the household. The sampling frame used is the master
ipling frame maintained by ACNielsen since 1966 when the basic sampling
imework was first acquired from the Department of Statistics, Government of
ilaysia (Jones, 1978). Since then the frame has been constantly updated, to keep
·east with the movements and growth of the population.

The frame was developed by dividing the geographical land areas of
ninsular Malaysia into enumeration blocks of approximately 200 to 240 dwelling
its (DUs). These enumeration blocks then form the primary sampling units
SUs) in the sample selection. Therefore, the sampling frame used for Media Index
rvey is a collection of 14,263 PSUs (ACNielsen sampling frame) covering the
al survey areas in Peninsular Malaysia.

Each PSU has detailed and clearly recognizable boundary particulars and
s. Thus, the frame is exhaustive, non-repetitive and traceable in the field. Every
J is classified into one of the 3 strata. The first stratum is the metropolitan towns
) consisting of 4 major cities that have been identified as central of all
keting activities. These 4 cities are Penang/Butterworth, Ipoh, Kuala
ıpur/Petaling Jaya and Johor Bharu.

The second stratum covers all other urban towns (UT) with population of
)00 and above. All the state capitals and other major cities fall into this stratum
ept the above 4 major cities that form the metropolitan towns. The last stratum is
rural areas (RT), covering areas with population below 10,000 and other
ıaining areas of Peninsular Malaysia. The MT stratum can be further sub-
tified into individual cities, while the UT and RT strata can also be sub-stratified
geographical region. The table below illustrates the definition of the 4
ıgraphical regions.

Table 3.1: Definition of Geographical Regions

Region	State group
North	Perlis
	Kedah
	Pulau Pinang
	Perak
Central	Selangor
	Wilayah Persekutuan Kuala Lumpur
	Negeri Sembilan
South	Melaka
	Johor
East	Kelantan
	Terengganu
	Pahang

These four regions are in fact quite different in many aspects. The ethnic position varies between regions. For example, the 1991 population census rted that about 90% of citizens in East region are Malays. In the other 3 ns, Malays make up between 48% to 57% of the population, while the Chinese e up about 35% of the population. The majority of Indians are living in the h and Central regions, making up 11% and 15% of the population in each region ectively. Besides, pronounced variations in spoken language (dialects) and iomic activities also exist between regions, even within the same state. eover, there are also different influencing factors such as in the South region ch the market behavior is very much driven by reactions from neighboring ;apore. In the Central region, the market behavior, social changes and elopment at large are determined at Kuala Lumpur / Petaling Jaya MT.

These differences have significantly impacted the social structure, market elopment and consumption habits of the population within each region. refore, in the sampling design of Media Index Survey, the selection process was ially performed separately by the 12 locations as shown in Table 3.2.

Table 3.2: Distribution of PSUs and Average Size of PSUs in The Frame and Sample by Stratum According to Geographical Regions

Stratum and Geographical region	Frame		Sample	
	No of PSUs	Average Size of PSU	No of PSUs selected	Average Size of PSU
etropolitan Town (MT)				
nang/Butterworth	788	222	29	230
oh	493	232	17	225
iala Lumpur/Petaling Jaya	2837	228	101	225
hor Bharu	738	224	16	215
ban Town (UT)				
orth	1190	222	39	224
entral	1127	226	32	228
uth	1027	220	29	205
ist	886	231	30	231
aral Area (RT)				
orth	1873	214	51	216
entral	850	220	21	221
uth	973	219	33	222
ast	1481	222	41	220
OTAL	**14263**	**223**	**439**	**222**

Table 3.3: Distribution of Dwelling Units in Frame and Sample and lumber of Completed Interviews by Stratum and by Geographical Regions

Stratum and Geographical region	Frame	Sample	Interviews
letropolitan Town (MT)			
enang/Butterworth	174690	6667	599
ooh	114284	3825	345
.uala Lumpur/Petaling Jaya	646002	22750	1719
ohor Bharu	165410	3442	322
Jrban Town (UT)			
lorth	264740	8730	818
.entral	255249	7062	670
iouth	226177	5947	628
last	204288	6922	682
tural Area (RT)			
lorth	401133	11010	1018
.entral	186832	4634	382
iouth	212826	7318	712
3ast	329227	9022	843
TOTAL	**3180858**	**97329**	**8738**

33

Selection Process

The units sampled at the first stage are the PSUs within each stratum and
on. Systematic selection was employed with the PSUs arranged in geographical
r from north to south. As each PSU is approximately equal in the number of
lling units (i.e. average 220 DUs), the PSUs were actually selected with
ability proportional to the number of DUs within each stratum as well as towns
in each stratum. This selection ensured the selection of a well spread and
esentative sample PSUs, that is also self weighting. The number of PSUs
cted is reflective of the population density in the areas.

Dwelling units within each sample PSU were the secondary sampling units
Us). The sample SSUs were selected proportionally to the number of DUs in the
pled PSUs using a random start number and a fixed sampling interval. For the
ropolitan towns and other urban towns, the sampling interval of 1 in 6 was
ployed with the random start number between 1 to 6 generated in the office for
h PSU. In rural stratum, a sampling interval of 1 in 7 was employed with the
dom start number between 1 to 7 generated for each PSU. Therefore, the
ction process was done with approximately overall uniform sampling fraction.

Subsequently, one respondent aged 15 years and above was selected
domly from every selected SSUs for interview. Besides information on
ividual media consumption habits, basic household data was also collected from
selected respondents.

Methods of Estimating Sampling Errors

The sampling error of a survey is usually measured in terms of the standard
or for a particular statistic such as mean or proportion of a characteristic. Since

34

sampling design for Media Index Survey involved features of stratification and

tering, the standard formulas for sampling errors found in statistical textbooks is

propriate because the formulas do not take into account these features. These

ıulas such as $\sqrt{pq/n}$ and $s/\sqrt{n}$ for the estimation of standard errors of a

ıortion and mean respectively, only apply for samples obtained by simple

ıom sampling. Therefore, the use of these formulas in Media Index Survey will

to understate the sampling errors of the survey estimates.

The computation of estimate sampling errors of survey estimates is done

g Taylor Expansion of ratio of two random variables. Let the ratio estimate, $r =$

, where y represents the total sample value for variable y, and x represents the

l number of cases in the group or subgroup under consideration. Therefore, r

be any percentages or means depend on the variable y.

The variance of r is computed using the formula as below with the standard

ır being the square root of the variance.

$$(r) = \frac{1-f}{x^2} \sum_{h=1}^{H} \frac{a_h}{a_h - 1} \left(\sum_{j=1}^{a_h} z_{hj}^2 - \frac{z_h^2}{a_h} \right)$$

ıre $z_{hj} = y_{hj} - rx_{hj}$ and $z_h = y_h - rx_h$.

h represents the stratum which varies from 1 to H,

a_h is the total number of PSUs selected in the hth stratum,

y_{hj} is the sum of the values of variable y in PSU j in the hth stratum,

x_{hj} is the sum of number of respondents in PSU j in the hth stratum,

f is the overall sampling fraction which is small and will be omitted in

calculation

One of the computer software for computation of standard errors that based

'aylor Expansion is CLUSTERS. However, the software is not available for this

y. MS Excel worksheet, in turn, is used in this study to compute the standard

rs for above formulation. The calculation process starts with obtaining the y_{hj}

x_{hj} for each of the sample PSU. Then, z_{hj} and z_h are calculated for each PSU and

um with appropriate input of y_{hj} and x_{hj}, and finally $var(r)$ is computed for

y stratum and total sample. The same calculation process is repeated for every

ables and domains under study. The calculation worksheet is attached in

)endix II.

Sampling errors for the Media Index Survey are calculated for selected

ables considered to be of primary interest. These variables include the

sehold variables and individual variables. The household variables consist of

mates of average number of cars, number of members aged 15 years old and

ve, number of members aged below 15 years old and number of television sets in

ousehold. Other household variables include estimates of household ownership

22 selected consumer durable goods such as telephone, refrigerator, air

iditioner and etc. Estimation of 14 household products purchased by the

pondents during the past one month will also be made.

Variables such as the number of days that respondents turned on television in

veek and the usual activities will be studied. Besides, estimates on newspaper

dership, banking services and literacy will also be examined. The list of selected

·iables for computation of sampling errors is in Table 3.4. The sampling errors

: computed by total sample, by stratum, by ethnic group, by gender and by age

)up.

Table 3.4: List of Selected Variables for Sampling errors, MIS

	Household Variable	Estimate	Base Population
	Number of cars	Mean	All household
A1	Durable item - portable radio/cassette player	Proportion	All household
A11	Durable item - camera	Proportion	All household
A12	Durable item - telephone	Proportion	All household
A13	Durable item - mobile phone	Proportion	All household
A23	Durable item - electric oven	Proportion	All household
A24	Durable item - microwave oven	Proportion	All household
A25	Durable item - electric blender/mixer/processor	Proportion	All household
A26	Durable item - cake mixer	Proportion	All household
A27	Durable item - electric rice cooker	Proportion	All household
A28	Durable item - toaster	Proportion	All household
A29	Durable item - refrigerator	Proportion	All household
A3	Durable item - compact disc player	Proportion	All household
A30	Durable item - washing machine	Proportion	All household
A31	Durable item - vacuum cleaner	Proportion	All household
A32	Durable item - water heater	Proportion	All household
A33	Durable item - portable fan	Proportion	All household
A34	Durable item - ceiling fan	Proportion	All household
A35	Durable item - air conditioner	Proportion	All household
A36	Durable item - sewing machine	Proportion	All household
A37	Durable item - personal computer	Proportion	All household
A4	Durable item - hi fi equipment	Proportion	All household
A5	Durable item - karaoke set	Proportion	All household
ABV15	Number of members aged 15 years and above	Mean	All household
BLW15	Number of members aged below 15 years	Mean	All household
IBUY1	Bought in past 1 month - instant noodles	Proportion	All household
IBUY13	Bought in past 1 month - shower gel	Proportion	All household
IBUY14	Bought in past 1 month - fabri conditioner	Proportion	All household
IBUY15	Bought in past 1 month - concentrated detergents for clothes	Proportion	All household
IBUY17	Bought in past 1 month - mineral water	Proportion	All household
IBUY18	Bought in past 1 month - baby diapers	Proportion	All household
IBUY19	Bought in past 1 month - energy drinks	Proportion	All household
IBUY20	Bought in past 1 month - corn oil	Proportion	All household
IBUY22	Bought in past 1 month - powdered condensed milk	Proportion	All household
IBUY3	Bought in past 1 month - breakfast cereals	Proportion	All household
IBUY4	Bought in past 1 month - UHT milk	Proportion	All household
IBUY5	Bought in past 1 month - instant coffee	Proportion	All household
IBUY6	Bought in past 1 month - tea bags	Proportion	All household
IBUY9	Bought in past 1 month - Hi calcium milk powder	Proportion	All household
SET	Number of TV sets	Mean	All household

Table 3.4: List of Selected Variables for Sampling errors, MIS

	Individual Variable	Estimate	Base Population
VT1	Activity-Participating in sports	Proportion	All adult aged 15+ years old
VT2	Activity-Reading newspaper and magazines	Proportion	All adult aged 15+ years old
VT4	Activity-travelling	Proportion	All adult aged 15+ years old
VT5	Activity-Visiting friends	Proportion	All adult aged 15+ years old
VT6	Activity-Listen to radio	Proportion	All adult aged 15+ years old
IK1	Banking facility - life insurance	Proportion	All adult aged 15+ years old
IK13	Banking facility - ATM card	Proportion	All adult aged 15+ years old
IK15	Banking facility - housing loan	Proportion	All adult aged 15+ years old
IK6	Banking facility - saving account - bank	Proportion	All adult aged 15+ years old
READ	Ability to speak BM	Proportion	Chinese, Indian and Others adult aged 15+ years old
TALK	Ability to Read BM	Proportion	Chinese, Indian and Others adult aged 15+ years old
Q1	Number of days watched TV1 in a week	Mean	All adult aged 15+ years old
Q2	Number of days watched TV2 in a week	Mean	All adult aged 15+ years old
Q3	Number of days watched TV3 in a week	Mean	All adult aged 15+ years old
ER1	English literate	Proportion	All adult aged 15+ years old
ER3	Bahasa Malaysia literate	Proportion	All adult aged 15+ years old
7DA	Read any dailies in past 7 days	Proportion	All adult aged 15+ years old
7DB	Read Bahasa daily in past 7 days	Proportion	All adult aged 15+ years old
7DE	Read English daily in past 7 days	Proportion	All adult aged 15+ years old
YTDA	Read any dailies yesterday	Proportion	All adult aged 15+ years old
YTDB	Read Bahasa daily read yesterday	Proportion	All adult aged 15+ years old
YTDE	Read English daily yesterday	Proportion	All adult aged 15+ years old

Design Effects and Coefficients of Intra-class Correlation

In addition to the standard errors, the design effect (*deft*) for each estimate is computed. The *deft* is defined as the ratio of the standard error of MIS pling design to the standard error assuming simple random sampling had been 1. A deft with value of 1.0 indicates that the sampling design is as efficient as a ple random sampling, while a value greater than 1.0 indicates the increased in the pling error due to the use of more complex sampling design.

The coefficient of intra-class correlation, *roh* measures the rate of iogeneity of PSUs. In practical survey clusters, *roh* tends to be greater than zero ause the populations are always not randomly distributed into groups and iters. Often, same type of dwelling units was built within the vicinity and likely ie occupied majority by certain ethnic group. Also, populations may be clustered some common characteristics within the groups.

The coefficient of intra-class correlation (*roh*) is calculated as follow:

$r = (deft^2 - 1)/(\bar{b} - 1)$ where $\bar{b}$ is the average number of elements per cluster.

Apart for the total sample and by stratum, *roh* is also be generated by iclass for all the selected variables. Table 3.5 gives the $\bar{b}$ value used for the *roh* culation corresponding to each geographic domain and subclass. The formulation $-n / a$ is used to calculate $\bar{b}$ value in Table 3.5.

Table 3.5: The $\bar{b}$ Used in Calculation of The *roh*

eographical Domains	$\bar{b}$	Race	$\bar{b}$	Gender	$\bar{b}$	Age Group	$\bar{b}$
tal Sample	19.9	Malay	13.5	Male	9.0	15-29	6.2
r	18.3	Chinese	7.7	Female	11.0	30-49	8.9
"	21.5	Indian/Others	3.8			50+	5.3
"	20.2						