

SAMPLING DESIGN FOR DOMESTIC TOURISM SURVEY, MALAYSIA

Norhayati Jantan and Hartini Yaacob¹

Abstrak

Kertas ini membentangkan metodologi pensampelan bagi Penyiasatan Pelancongan Domestik (PPD) yang dilaksanakan secara bulanan oleh Jabatan Perangkaan Malaysia (JP). Objektif penyiasatan ini adalah menyediakan maklumat berkenaan profil dan ciri-ciri pelawat domestik untuk sesuatu tahun tertentu. Maklumat penyiasatan ini juga turut digunakan sebagai input bagi penyusunan Akaun Satelit Pelancongan dan juga analisis lain oleh pengguna data.

PPD meliputi semua negeri di Malaysia dan distratakan mengikut bandar dan luar bandar. Rekabentuk pensampelan yang digunakan dalam menentukan saiz sampel adalah Pensampelan Berstrata Dua Peringkat. Pemilihan sampel peringkat pertama (PSU) telah dibuat dengan menggunakan Kaedah "Probability Proportionate to Size" (PPS) manakala pemilihan sampel peringkat kedua (SSU) menggunakan Kaedah Pensampelan Bersistematik.

Kata Kunci : Metodologi Pensampelan, Pensampelan Berstrata Dua Peringkat

Abstract

This paper presents the sampling design used in the Domestic Tourism Survey (DTS) carried out monthly by the Department of Statistics, Malaysia (DOSM). The objective of this survey is to provide data on the profile and characteristics of domestic tourists in a particular year. Data from this survey also serve as an input to the compilation of Tourism Satellite Accounts (TSA) and for analysis by users.

DTS covers all states in Malaysia and is stratified by urban and rural. Two-stage stratified probability sampling is used in designing and determining the sample size. The primary sampling units (PSUs) were selected by using the Probability Proportionate to Size (PPS) method, while the secondary sampling units (SSUs) were selected using Systematic Sampling Method.

Keywords: Sampling design, Two-stage stratified probability sampling

¹Norhayati binti Jantan is currently the Director, Methodology and Research Division and Hartini binti Yaacob is Principal Assistant Director of the same Division.

Introduction

Domestic tourism is the tourism of resident visitors within the economic territory of the country of reference (World Tourism Organization, 2001). The purposes of travel include recreational, leisure or business. Household concept is used in compiling Domestic Tourism Survey (DTS) data. It covers several areas related to travel and expenditure patterns, among which are; detailed monthly household spending on domestic tourism, types of trips, main purpose of the trips, main activities during the trips, demographic profile of domestic tourists and income profile of domestic tourists.

DTS is the first national household survey dedicated to tourism which is conducted by DOSM since 2007. The information collected in this survey serve as an input to the effective and efficient development and formulation of tourism policy and programmes. It provides policy-makers with the latest reliable data for decision-making on domestic tourism demand. Specifically, the survey aims to provide inputs for the compilation of the Tourism Satellite Accounts (TSA) by the Department.

The variable of selection for this survey is expenditure on tourism activities by household. However prior to 2007, no data were collected for this variable in the household surveys conducted by the Department. Therefore, expenditure on Package Tours and Accommodation Services from the 2004/2005 Household Expenditure Survey was used as a proxy. The samples were selected to represent urban and rural areas of each state in Malaysia in order to enable in-depth analysis at state and stratum level. DTS was carried out using the personal interview approach which involved households in the living quarters (LQs) in the selected enumeration blocks (EBs). About 3,800 EBs were selected, which covered around 30,000 LQs throughout Malaysia. Approximately 2,600 LQs will be visited every month.

Literature Review

Sample designs for household surveys in developing and transition countries have many common features. Most of the surveys are based on multistage stratified area probability sample designs. These designs are used primarily for frame development and for clustering interviews in order to reduce cost (United Nations, 2005). Kish (1965) suggests that sample design covers two domains: a selection process, where the rules, standards and procedures are developed to determine the sampling frame and survey population and an estimation process, where sample statistics can be used in determining satisfactory sample sizes. Further, Kish states that a good sample design requires the judicious balancing of four broad criteria: goal orientation, measurability, practicality and economy. Goal orientation is meant to consider both the selection process and the estimation process in terms of meeting research objectives. The research objective for the DTS is to determine travel behaviour patterns of a cross-section of Malaysian residents. The research must address detailed characteristics of the household, each person in the household and the trips each person makes on a pre-selected day.

South Africa designed DTS to be representative at provincial level and within provinces at metro/non-metro level. Within the metros, the sample is further distributed according to geographical type. The four geographical types are urban formal, urban informal, farms and tribal. The sample was based on a stratified two-stage design with Probability Proportionate to Size (PPS) sampling of primary sampling units (PSUs) in the first stage and sampling of dwelling units (DUs) with systematic sampling in the second stage (Statistics South Africa, 2009).

In developing countries like India, such sampling frame is neither readily available nor can it be easily prepared since developing new frame is an expensive proposition. The survey design adopted a three stage stratification in which a ready-made frame could be used at least for the first two stages and a sampling frame developed in the last stage (World Tourism Organization Publication).

Jordan is using the same sample design of the Household Income and Expenditure Survey 2008 because it is a relatively large size sample and provides a high level of accuracy data. A two- phase stratified cluster sampling method was used in designing the survey sample. In the first stage, a sample of clusters was drawn using the sampling method that commensurate with the size and drawn regularly by treating the number of households in each cluster as the weight of that cluster, while the second stage included the selection of the targeted households sample in each cluster that has a size of 8 households (Department of Statistics, Jordan, 2008).

Face to face structured interviews were carried out in all the 13 political regions in Namibia. A total sample of 1,200 households was drawn countrywide. The sample was divided into what is called political regions as stipulated by the Central Bureau of Statistics (CBS). Sixty samples of eight PSUs were interviewed countrywide (both rural and urban areas) and the number of PSUs per region differed depending on the population size. In each PSU, 20 households were selected for an interview and systematic random sampling method was employed. The sample data was weighted according to the total number of households per rural and urban areas and region. (Namibia Tourism Board, 2008).

Among the countries which conducted similar survey are: New Zealand (New Zealand's Domestic Travel Survey), Australia (The National Visitor Survey), United Kingdom (United Kingdom Tourism Survey), China (Domestic Tourism Sample Survey), Philippines (Domestic Visitors Survey), Turkey (Household Domestic Tourism Survey), Canada (Canadian Travel Survey), India (Domestic Tourism Survey), Thailand (Domestic Tourism Survey) and Vietnam (Tourism Expenditure Survey).

Concepts and Definitions

The concepts and definitions used are adopted from DTS publication which is in line with International Recommendations for Tourism Statistics 2008 by United Nations and World Tourism Organization (UNWTO).

Domestic visitor

A domestic visitor is defined as a resident or those living in Malaysia for at least a year including expatriates and non-citizens who take a trip to main destinations outside his/her usual environment. The length of the trip should take less than a year for the purpose of business, leisure or personal other than to be employed by a resident entity in the place visited. A domestic visitor is categorised into:

- i. Domestic tourists - referring to those who travel in Malaysia for at least one night (24 hours); and
- ii. Excursionists or daily visitors - referring to those who travel in Malaysia for less than 24 hours.

Tourism trip

A tourism trip refers to travel that takes the traveller to a main destination outside his/her usual environment, for less than a year, for any main purpose (business, leisure or other personal) other than to be employed by a resident entity in the place visited.

A domestic tourism trip refers to travelling to a main destination within the country of residence of the visitor. This can be divided into two:

- i. Same day trip/daily trip
 - a) Travels out of the house and cover a distance of 50 km to and fro, or more, and takes four hours and more; and
 - b) For distances less than 50 km to and fro, the time taken has to be four hours and more and has to use tourism facilities such as transportation/accommodation/food and drinks/recreation.
- ii. Overnight trip
Refers to travel that requires a person to leave home at least for one night (24 hours) with the purpose of travel relating to tourism and using tourism facilities. The duration of travel must be less than 12 months.

Usual residence

The place of usual residence is the geographical place where the enumerated person usually resides.

Domestic tourism expenditure

Domestic tourism expenditure refers to any kind of expenses spent by a resident visitor within the economy of reference. Tourism expenditure refers to the amount paid for the acquisition of consumption goods and services, as well as valuables, for own use or to give away, for and during tourism trips. It includes expenditure by visitors themselves, as well as expenses that are paid for or reimbursed by others.

Scope and Coverage

The target population of DTS is to cover all residents who stayed at least one year in Malaysia including non-citizens and expatriates. It covers both urban and rural areas in the 16 states in Malaysia. They are household members who live in private living quarters, have their own personal life and are known as usual members. Usual members are persons who regard the dwelling as their own or main home. Others present are considered to be visitors and thus are not covered in the survey.

For sampling purposes, certain areas are excluded due to inaccessibility and safety reasons such as high risk enumeration blocks, inaccessible and remote areas where time and cost are most required. Besides that, the persons who live in the institutional living quarters such as hotels, hospitals, nursing homes, welfare homes and prisons are excluded.

The exclusion of institutional living quarters is to avoid duplication of persons or double counting during enumeration. These persons' information will be reported by the selected households based on the concept of usual members. For inaccessible and remote areas, long distances and uneven paths would incur higher cost for logistics and time consuming.

Sampling Frame

The sampling frame for household surveys is maintained under the Household Frame (HH Frame). HH Frame consists of information on enumeration blocks such as state, administrative districts, strata, number of LQs and list of building units in each EB. Building units are further categorised into LQs, institutional LQs and non-LQs. For sampling purposes, the number of occupied LQs will be used as the measure of EBs size for applying PPS method.

For DTS 2007 until 2011, sampling frame is derived from the 2000 Population and Housing Census (Census 2000). The household frame is made up of EBs which is geographically contiguous areas of land with identifiable boundaries. On average, each EB contains about 80–120 LQs. Generally, all EBs are formed within gazetted boundaries, i.e. administrative district, *mukim* or local authorities' areas.

Urban areas for Census 2000 were defined as gazetted areas with their adjoining built-up areas which had a combined population of 10,000 or more. Built-up areas were defined as areas contiguous to a gazetted area and has at least 60 per cent of their population (aged 10 years and over) engaged in non-agricultural activities as well as having modern toilet facilities in their housing units.

The sampling frame for DTS 2012 onwards is derived from the Population and Housing Census 2010. The urban areas are now defined as gazetted areas with their adjoining built-up areas with combined population of 10,000 or more and built up areas defined as areas contiguous to a gazetted area and had at least 60 per cent of their population (aged 15 years and over) engaged in non-agricultural activities.

The definition of urban areas also takes into account the special development area namely the development area which is not gazetted and can be identified and separated from the gazetted area or built-up area of more than 5 km and a population of at least 10,000 with 60 per cent of the population (aged 15 years and over) involved in non-agricultural activities.

The classification of areas by stratum is as follows:

Stratum	Population of Gazetted and Built –up Areas
i. Metropolitan	75,000 and above
ii. Urban large	10,000 to 74,999
iii. Urban small	1,000 to 9,999
iv. Rural	less than 1,000

For sampling purposes, the above broad classifications were found to be adequate for all states in Peninsular Malaysia and the Federal Territories. However, for Sabah and Sarawak, due to difficulties in accessibility, the rural stratum has to be further sub stratified based on the time taken to reach the area from the nearest urban centre.

For the purpose of urban and rural analysis, the four stratum used in the classification of areas by stratum are redefined as follows:

Urban = Metropolitan + Urban Large
Rural = Urban Small + Rural

Sampling Process

Sampling is the process of selecting and observing a part (sample) of the population in order to make inferences about the whole population (Kish, 1965). In performing the sampling design for DTS, several inputs would be required such as previous survey data and level of analysis. Besides, indication of important variables or expected

findings of DTS is required to determine the appropriate variable of selection. This variable will be used to estimate the sample size, for example, mean expenditure per trip or number of trips. Other factors such as cost and time will be taken into account in determining the sample size.

For DTS, the probability sampling method is applied in order to obtain the estimates for population parameters regarding domestic tourism even though it will incur higher cost and time as compared to non-probability sampling. In probability sampling, each unit in the sampling frame will have its known probability to be selected to represent the whole population. Besides that, by using probability sampling, the sampling error of estimates can be calculated in which later be used as a guide for assessing the survey data results.

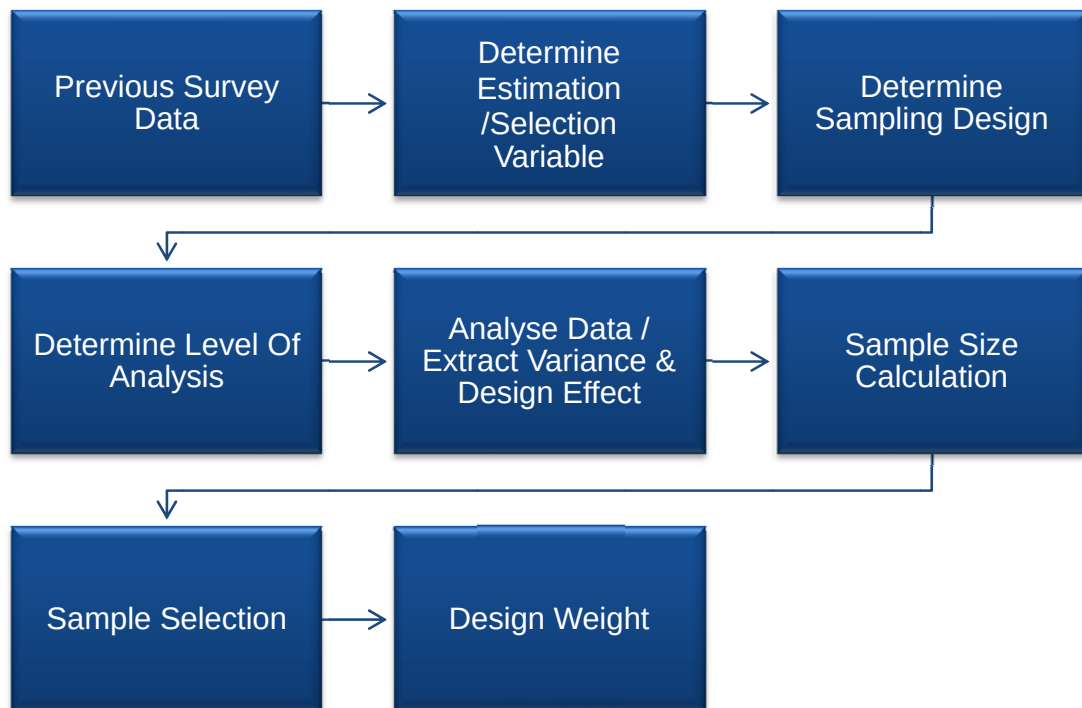
There are several different methods namely simple random sampling (SRS), stratified sampling and cluster sampling in which a probability sampling can be employed. According to Lohr, SRS is the simplest form of probability sample where every possible unit in the population has the same chance of being the sample. In a stratified sampling, the population is divided into subgroups called strata. Then an SRS is selected from each stratum, and SRSs in the strata are selected independently. Elements in the same strata tend to be more similar. Kish stated that cluster sampling denotes methods of selection in which the sampling unit, the unit of selection, contain more than one population element; hence the sampling unit is a cluster of elements. Each element must be uniquely identified with one and only one sampling unit.

The method chosen depends on a number of factors, such as the available sampling frame, cost of the survey and level of analysis. When considering a good probability sample design, the goal should be to minimise the sampling error of the estimates for the most important survey variables, while minimising the time and cost of conducting the survey.

Diagram 1 illustrates the probability sampling procedure to determine the sample size for DTS. Previous survey data or any relevant proxy data are required in every sampling process to understand the nature and structure of the data before any sampling techniques can be applied. Usually there are a few potential variables in the data that can be used for sample size estimation. The selection of the most appropriate variable would generate optimum sample size. For this study, expenditure by the residents on tourism activities is used as the variable of selection.

The other activities in the sampling process include determining the level of analysis, analysing variance and design effect of the previous results, estimating sample size, sample selection and calculation of sampling weight for estimation.

Diagram 1: Sampling Process of Probability Sampling



Sampling Design

Sampling design for DTS is based on multistage stratified area probability sample design, namely two-stage stratified probability sampling. This design is used due to the constraints in household frame where only list of LQs are available. The other reason for using this design is because the population is heterogeneous, or dissimilar, where certain homogeneous or similar sub-populations can be isolated (strata). Some advantages of using stratified sampling over simple random sampling are:

- the cost per observation in the survey may be reduced;
- estimates of the population parameters may be needed for each sub-population; and
- increased accuracy at given cost.

a) Stratification

To cater to the stratification of the frame is based on geographical disaggregation. The main stratum consists of all states in Malaysia including Wilayah Persekutuan (WP) Kuala Lumpur, Wilayah Persekutuan (WP) Labuan and Wilayah Persekutuan (WP) Putrajaya. The second stratum is the urban-rural dichotomy, created within their

respective states, with exception of WP Kuala Lumpur and WP Putrajaya (urban only). The stratification is as follows;

- Main Stratum – All states in Malaysia
- Second Stratum – Urban and rural of each state

b) Sample Estimation

Analysis of the DTS is done at state and strata level, accordingly the calculation of sample size is by state and strata. There are several factors that influence the determination of the sample size. One of them is the level of analysis required whereby the more disaggregated the analysis, the sample size has to be large enough to cover the diversity of the population. However, the bigger sample size will require more resources such as budget, staff and time.

Besides level of analysis, other factors that determine the sample size are mean and variance of the variable of selection, design effect, margin of error, confidence level, cost and time of the survey.

The formula used for calculating the sample size is as follows:

$$n_{0j} = \frac{z^2 s^2}{d^2}; j = 1, 2, 3, \dots, k$$

Where,

z is the confidence level (1.96 for 95% confidence level)

d is the margin of error

s is the standard deviation (std.dev) of variable of selection

Step 1

The margin of error 'd' measures the precision of an estimate from a single sample approximates the population value. The margin of error will get narrower as the sample size increases. The margin of error selected depends on the precision needed to make population estimates from a sample and is a controlled factor in the sample size estimation in obtaining desired precision. Margin of error is derived as follows:

$$d = RSE \cdot \bar{y}$$

Where,

RSE = Targeted relative standard error

$\bar{y}$ = mean expenditure of previous DTS

RSE is targeted to be not more than 10 per cent at urban and rural areas of every state and $\bar{y}$ is the mean expenditure of previous DTS. An example of how the margin of error is derived is given for State A:

State	Strata	Mean, $\bar{y}$	RSE	d
A	Urban	242.92	0.1	24.29
	Rural	195.01	0.1	19.50

Step 2

After extracting the value of margin of error and standard deviation, calculation of the sample is done with 95 per cent confidence level as another control factor in the sample size estimation. Confidence level is the estimated probability that a population estimate lies within a given margin of error. The calculation to obtain initial sample size, n_0 is given below:

State	Strata	Mean, $\bar{y}$	RSE	d	std.dev	n_0
A	Urban	242.92	0.1	24.29	1084.44	7,656.10
	Rural	195.01	0.1	19.50	452.64	2,069.67

If the standard deviation of the data is large, a larger sample size is required to overcome the diversity of the data. A large variance will result in larger sample sizes. Sometimes size of samples is likely to increase largely due to outliers. As outliers will possibly occur in any set of data, outlier analysis and treatment have to be done not only for data analysis purposes but also for sampling process to obtain optimum and representative samples.

Step 3

The initial sample size n_0 , then collaborates with the Finite Population Correction (FPC) factor as follows:

$$n_{oj} = \frac{n_0 j}{1 + \frac{n_0 j}{N_j}}; j = 1, 2, 3, \dots, k$$

where,

n_{oj} is the minimum sample size for j^{th} sub-population, derived from SRS

N_j is the total for j^{th} sub-population

k is number of sub-population

FPC is routinely used in calculating sample sizes using simple random sampling. It has very little effect on the sample size when the sample is small relative to the population but it is important to apply the FPC when the sample is large (10 per cent or more) relative to the population (Cochran, 1987). After taking into consideration the size of the population, sample size is given as n_1 :

State	Strata	Mean, $\bar{y}$	RSE	d	std.dev	n_0	n_1
A	Urban	242.92	0.1	24.29	1084.44	7,656.10	7,629.98
	Rural	195.01	0.1	19.50	452.64	2,069.67	2,065.89

Step 4

Sample size calculations were based on the assumption of simple random sampling. Therefore, sample designs other than simple random sampling have an impact, called design effects (DE) on sampling variability. As a result of this impact, DE is an important consideration when determining sample size.

Therefore, the sample size will be adjusted with the DE and calculated as follows:

$$n_{2j} = n_{1j} \times DE$$

Where,

$$DE = \frac{\text{Variance estimate from complex design}}{\text{Variance estimate from SRS}}$$

DE measures the effect of the design on the variance of an estimate. It can also be interpreted as the loss of effectiveness by the use of cluster sampling, as compared to simple random sampling. Sample size after applying the DE for each stratum are as follows:

State	Strata	n_0	n_1	DE	n_2
A	Urban	7,656.10	7,629.98	2.24	17,091.16
	Rural	2,069.67	2,065.89	2.00	4,131.78

Design effect of 1 indicates that the precision gained using complex design is the same when the simple random sampling is used. Thus, same sample size is needed to obtain the same sampling precision found with a simple random sample. A design effect which is greater than 1 indicates that more cases would be needed; alternately, fewer cases

would be needed to obtain the same sampling precision found with a simple random sample if the design effect is less than 1.

Step 5

Due to non-response problem such as refused to response and no one at home the, sample size must take into account the non-response rate and is adjusted as follows;

$$n_{3j} = n_{2j} \times \frac{1}{\text{Response Rate}}$$

Response rate of the previous DTS were used to estimate the current sample size. Response rates are different for all states and stratum. The lower the response rate, a higher sample size should be taken. Response rates for DTS lies between 85 to 95 per cent. For instant, response rate of urban and Johor are 87 and 92 per cent respectively, thus n_3 is given by;

State	Strata	n_0	n_1	n_2	n_3
A	Urban	7,656.10	7,629.98	17,091.16	19,645
	Rural	2,069.67	2,065.89	4,131.78	4,491

n_3 is a number of individual that will be sample, however LQs is the smallest sampling unit available in HH Frame. To determine number of LQs, information on household size is needed where the number of individual is divided by the household size as follows;

$$nj = n_{3j} \times \frac{1}{\text{Household Size}}$$

When household size for urban and rural, Johor are 3.7 and 5.3 respectively, number of LQs samples estimated as;

State	Strata	n_3 , individual	n , LQs
A	Urban	19,645	5,309
	Rural	4,491	847

Therefore, sample size of LQs can be finalised as;

$$n = \sum_{j=1}^k n_{3j} \quad j = 1, 2, \dots, k$$

Step 6

The following step is to determine the number of EBs to be sampled. With the assumptions that eight LQs will be selected per single EB, the number of EBs and LQs is given as follows;

State	Strata	LQs	EBs
A	Urban	5,309	664
	Rural	847	106

Besides the factors mentioned above, there are indirect factors that influence the sample size. These factors are cost and time of data collection, processing and dissemination. Adjustments will be made to the calculated sample size in relation to the cost and time of the whole survey. This adjustment is based on the RSE of the previous survey to maintain reliability of estimates.

Hence, the total annual sample size for DTS 2010 to 2012 by urban and rural is as follows:

Table 1: Sample Size for DTS 2010-2012 by Strata

Year	Urban		Rural		Malaysia	
	EBs	LQs	EBs	LQs	EBs	LQs
2010	2,162	17,296	1,638	13,104	3,800	30,400
2011	2,367	18,936	1,508	12,064	3,875	31,000
2012	2,388	18,104	1,421	11,368	3,809	29,472

c) Sample Selection

The sample for EBs is selected from each urban-rural stratum within the respective states. This process uses PPS method and is widely used by most of the developing and transition countries due to variability of measures of size. The measures of size generally takes into account the number of occupied LQs, households or persons in the PSUs based on the last population census. They may be significantly out of date and they may be markedly different from the current size because of such factors like growth in urban areas and shrinkage in other areas as a result of migration and natural disasters. Inaccurate measures of size lead to lack of control over the distribution of second stage units and this can cause serious problems in subsequent field operations. One of the solutions to overcome this problem of inaccurate measures of size is to conduct a new listing of LQs for the EBs that experienced major changes in the number of occupied LQs.

The selection of LQs as a secondary sampling unit is done by systematic method where all units have equal chances of selection. Currently, eight LQs will be selected within each EB. Nevertheless, in some cases the number of LQs can be more or less than eight depending on the latest information of the EBs. The formula for determining the secondary sample is as follows:

$$\text{Number of sample LQs for one EB} = \frac{\text{Latest number of LQs}}{\text{Number of existing LQs}} \times b$$

Where b = sample size for secondary sampling unit

Selection of LQs have to be carried out only for occupied LQs. Hence, the updating exercise to eliminate non response errors due to frame defects such as non-occupied LQs, LQs under renovation and non-LQs have to be conducted regularly.

For selected LQs, all members aged 15 years and above in the household will be interviewed. Each LQ may have a single or multiple households, depending on the arrangement made for preparation of food and basic expenditure of their members.

d) Sampling Weight

The sampling weight is applied to the survey data to generate population estimation. The calculation of weights is derived from the number of households selected as a sample minus the non response cases.

Data collection is done monthly and the data is analysed on a monthly, quarterly as well as yearly basis. Different types of weights are produced to cater for these periodical analyses where it is being prepared in two stages which are household weights and individual weights.

The formula for the household weights for three periods is as follows:

i) Monthly Weight (W_m)

$$W'_{msst} = N_{sst}/n'_{msst}$$

Where,

m = month 1 to 12

s = state 1 to 16

st = strata 1 to 2

n'_{msst} = responded sample for m-month, s-state and st-strata

N_{sst} = population for s state and st-strata

W'_{msst} = weight for m-month, s-state and st-strata

ii) Quarterly Weight (W_q)

$$W'_{qsst} = N_{sst}/n'_{qsst}$$

Where,

q	= quarter 1 to 4
s	= State 1 to 16
st	= strata 1 to 2
n'_{qsst}	= responded sample for q-quarter, s-state and st-strata
N_{sst}	= population for s state and st-strata
W'_{qsst}	= weight for q-quarter, s-state and st-strata

iii) Yearly Weight (W)

$$W'_{sst} = N_{sst}/n'_{sst}$$

Where,

s	= State 1 to 16
st	= strata 1 to 2
n'_{sst}	= responded sample for s-state st-strata
N_{sst}	= population for s state st-strata
W'_{sst}	= weight for s-state st-strata

After obtaining the weighted estimate of population by using household weights (labelled as w_1 , the combined ratio estimated method (household weight and population factor) was used to obtain the estimate of a specific characteristic in the survey population;

$$E' = \sum_{n=1}^k E_n \left(\frac{W_n}{w_n} \right)$$

E_n is the weighted estimate from the sample count of persons of a characteristic (e.g. domestic visitors) in a particular age-sex-ethnic-state (e.g. Malay males in the 15-19 age group in Johor).

W_n is the weighted estimate from the sample count of the number of persons in a particular age-sex-ethnic-state.

W_n is the independent mid-year population estimate of the number of persons in a particular age-sex-ethnic-state.

$\frac{W_n}{w_n}$ is the adjusted estimate of a characteristic in a particular age-sex-ethnic-state obtained by the age-sex-ethnic-state adjustment.

E' is obtained by adding the adjusted estimate for all age-sex-ethnic-state. Thus, is also $E_n \left(\frac{W_n}{w_n} \right)$ known as the ratio estimator of the characteristic.

Limitations

Statistical unit in the DTS is a household member where the important requirement is to have a list of individuals. Individual frame is not available in DOSM and for this reason, lists of EBs and LQs are used as a sampling frame. To obtain the estimate at population level, both sampling weights which are household weights and population factor are applied.

Non-coverage for certain areas is also a major concern for household surveys due to inaccessibility issue. As discussed above, inaccessibility areas were excluded in the DTS due to geographical constraints and safety reasons. As a result, some sampling units have no probability of inclusion in the sample. However, sampling weights were used for estimation which can represent these excluded areas.

Moving forward

Current stratification of urban and rural is used due to homogeneity in the characteristic within stratum. However, as the demographic and economic characteristics changes over time, the profiles of population especially in urban areas are no longer homogeneous. The demand for further stratification so call sub-strata within urban areas is needed to maintain the representativeness of the samples selected in the more vibrant states of W.P Putrajaya, W.P Kuala Lumpur, Selangor, Johor and Pulau Pinang.

Besides that, study is carried out to determine whether the current number of LQs per EB is still relevant to certain urban areas as the characteristics within the EB no longer follow the characteristics of strata but more towards the characteristics of cluster. The finding of the study suggests that certain urban areas in the more developed states need to be increased to cater to the different characteristics of households in the same

EB. However, the implementation of this suggestion can be applied after few similar studies for different types of survey are completed.

Conclusion

The sample size of DTS is influenced by several factors such as margin of error, confidence level, variance of the desired variable, total population, quality of the previous data, design effect and indirect impact by the desired cost constraints and time. Among these factors, quality of the previous data which contribute to variance estimation as well as design effect are the main considerations in determining sample size for DTS. Domestic tourism expenditure data is volatile in nature since it is not a common expenditure as compared to basic consumption expenditure. As a result, sample size is targeted at 10 per cent RSE.

References

- Cochran, W.G. (1987). *Sampling Techniques*. John Wiley and Sons, New York.
- Department of Statistics, Malaysia. (2010). *Domestic Tourism Survey*, Putrajaya: Department of Statistics, Malaysia.
- Department of Statistics, Malaysia. (2011). *Domestic Tourism Survey*, Putrajaya: Department of Statistics, Malaysia.
- Domestic Tourism Survey 2008. (2009). Statistics South Africa, South Africa.
- Domestics Tourism Survey 2008. (2009), Department of Statistics, Jordan.
- Household Sample Surveys in Developing and Transition Countries. (2005). United Nations New York.
- Jessen, R.J. (1978). *Statistical Survey Techniques*. John Wiley and Sons, New York.
- Kazmier, Leonard J. (2004). *Business Statistics*. McGraw-Hill, New York
- Kish, L. (1965). *Survey Sampling*. John Wiley & Sons, Inc., New York.
- Lam, Swee Kin. (1986). *Statistik STPM*, 3th.Ed. Syarikat United Book Sdn Bhd. Fajar Bakti Sdn Bhd.
- Lohr, S.L. (1999). *Sampling: Design and Analysis*. Duxbury Press, Pacific Grove, CA.
- Preliminary Domestic Tourism Survey Findings (2008), Namibia Tourism Board. Retrieved from <http://www.namibiatourism.com.na/>
- Stuart, A. (1994). *The Ideas of Sampling*. MacMillan Publishing Company, New York.
- The Indian Domestic Tourism Survey. (2002). Ministry of Tourism and Culture, India.
- Thompson, S. (1992). *Sampling*. John Wiley & Sons, New York.
- UNSD, United Nations Statistics Division, WTO, World Tourism Organization, OECD, Organisation for Economic Cooperation and Development and Eurostat, Statistical Office of the European Communities (2001). *Tourism Satellite Account (TSA): Recommended Methodological Framework*.
- United Nations and World Tourism Organization. (2010). *Recommendations on Tourism Statistics 2008*. United Nations publication, New York.
- Zainal Mat Saat. (1985). *Pengantar Statistik*, Penerbit Fajar Bakti Sdn Bhd.