

# Optimum Sampling Strategy for National Households in Abu Dhabi: With Implementation on the Household Master Sample

Dr. Mohammed Al Rifai  
Statistics Centre – Abu Dhabi, P.O. Box: 6036 Abu Dhabi, UAE

## Abstract

In Abu Dhabi, households are classified as “national households” where the head of household is citizen; “non-national households” family households with household head of other nationalities; and “shared households”, which are unrelated persons living independently in same housing unit. Although the national households represent 21.5% of the total number of households in Abu Dhabi, they are the most important domain for the analysis. For this reason, a special sampling strategy is required to increase the number of national households in the sample of the surveys. Based on the 2011 Abu Dhabi population census, Statistics Centre – Abu Dhabi (SCAD) constructed a household frame. A master sample of housing units was developed to support a stratified, two-stage sample design for future surveys, with stratification of enumeration areas. The enumeration areas (EAs) were stratified by the percentage of national households. In addition, a sample allocation criterion was devised to select the sample of the EAs to come up with the optimum total sample of national households in Master Sample 2012. This criterion assigned different weights to the strata based on total nationals in the sampling frame. The proposed sampling strategy produced significant improvements in the percentage of selected sample that were national households.

**Keywords:** Sampling frame, sample allocation, stratification, Master Sample

**DOI:** 10.7176/JESD/10-24-10

**Publication date:** December 31<sup>st</sup> 2019

## 1. Introduction

### 1.1 Abu Dhabi

Abu Dhabi emirate is the federal capital of the United Arab Emirates (UAE) and the largest of the seven emirates. Abu Dhabi emirate consists of three main regions, Abu Dhabi Region, Al Ain region and Al Dhafra region.

Throughout the last few decades, Abu Dhabi Government devoted outstanding efforts, to promote the emirate’s profile and reputation fledged by the most up-to-date systems, on all levels (economic, social and excellence in government performance,) and laid down a number of plans and initiatives, ensuring an appropriate and effective alignment of development gains/ benefits, and build a sustainable economy serving all the Emirate’s regions.

Statistics Centre – Abu Dhabi (SCAD) was established in accordance with Law #7 for the year 2008 as the main authorized body concerned with official statistics in the Emirate of Abu Dhabi. SCAD is responsible for the collection, classification, storage, analysis and dissemination of official statistics covering social, demographic, economic, environmental and cultural indicators.

### 1.2 How to improve Optimum sample for national households

The Statistics Centre – Abu Dhabi (SCAD) has an essential role to provide statistics to support strategies, initiatives, and plans that focus on the development of Abu Dhabi Emirate. As a result, SCAD conducts various sample household surveys to provide precise and highly accurate statistical indicators, with more emphasis on the nationals as a main target population of the Emirate’s development.

The use of a master sample is an effective approach for selecting the sample households for the various surveys to ensure accurate data is available to meet the government’s needs. SCAD is already developing plans for specific household surveys. One of the most important household surveys that is conducted on a regular basis is the Labor Force Survey (LFS). For efficiency, SCAD has multi-purpose modules attached to the LFS, covering special topics such as immigration, disability and education.

Now that the data and cartography from the 2011 Abu Dhabi Census of Population are available to provide an updated sampling frame for the household surveys in the Emirate, a master sample design is proposed to provide efficient samples for the household survey program of SCAD over the next 5 years or more. An important consideration for the master sample design is that the households in the sampling frame are classified into national households; non-national households, consisting of families with other nationalities; and shared households, which are collective households with unrelated persons living independently in a housing unit. Although the national households only represent 21.5% of the total number of households in the Abu Dhabi Emirate, they are a very important domain for the analysis in all future household surveys.

In order to propose an efficient sampling strategy that increases the number of national households in the master sample, the following questions needed to be answered. “What is the stratification method to stratify the enumeration areas in the sampling frame according to a nationality composition measure?” and “What is the

recommended method for allocating the sample size for those proposed strata?"

## 2. The Structure of Abu Dhabi Sampling Frame

The sampling frame for the Abu Dhabi master sample was based on the data and cartography from the 2011 Abu Dhabi Population Census. Most of the household surveys in the Abu Dhabi Emirate will follow a stratified two-stage sample design. The primary sampling units (PSUs) selected at the first stage for the master sample are the census enumeration areas (EAs), which are small operational areas defined on maps for the 2011 Census enumeration. The EAs have an average of about 120 households each, which is an effective size for conducting a new listing of households and controlling the fieldwork in sample clusters. There are 1883 EAs in the 2011 Abu Dhabi Census frame.

Abu Dhabi Emirate is divided into three administrative regions, Abu Dhabi, Al Ain and Al Dhafra. Each region is divided into districts, which are further divided into sectors. Each district was previously designated as urban or rural, although this was not taken into consideration in the 2011 Census. For this reason, the urban and rural categories will not be domains of analysis for the household surveys, but this classification can be used as one of the criteria for implicit stratification of the master sampling frame.

Another important consideration for the master sampling frame is the classification of the residential households according to the census definitions and concepts. The households are classified as national, non-national, and collective households. The distribution of households in the 2011 Abu Dhabi Census frame by type of household and region is presented in the following table.

**Table 1.** Distribution of Households in 2011 Household Frame by Region and Type of Household.

Region	Total No. EAs	% Total HHs	% National HHs	%Non- National HHs	% Shared HHs
Abu Dhabi	1215	64.4	58	70.2	52.8
Al Ain	551	29.2	37	23.6	38.4
Al Dhafra	117	6.4	5	6.2	8.7
Total	1883	100	100	100	100

## 3. Master Sample Design Approach for Household Surveys

The most cost-effective and efficient approach for providing the samples for household surveys over a period of years is through a master sampling frame, as indicated by Anthony G. Turner. By continually updating the cartography and data in a master sample of areas, the sample for each household survey remains representative over time. A master sample design is normally developed following a census of population and housing.

A master sample design generally involves the selection of a sample of primary sampling units (PSUs) of a sufficient size to provide sample dwelling units for each household survey over a period of several years. An individual survey can be based on all of the PSUs in the master sample, or on a subsample.

There are several advantages for using a master sample for the household survey program compared to selecting an independent sample for each survey. A major advantage is the cost savings. A master sample will reduce the overhead costs for each survey because the cost of compiling the frame, selecting the sample PSUs and obtaining the census maps for sample areas will be shared across the different surveys. The updated listing of households in sample PSUs can be shared for the different surveys, and updated each year. It is possible to invest in a cartographic update of master sample PSUs to benefit all household surveys.

The master sample can be used for many years, and the frame of households within the master sample of PSUs can be updated over time. There will be more consistency in the results from the household surveys with multiple rounds and for different surveys based on the master sample, since there will be more correlation in the socioeconomic characteristics of the sample for the different surveys.

The master sample provides flexibility in the sample size and allocation for different types of household surveys based on the master sample. The sample size for each survey may vary depending on the objectives of the survey. The master sample will include replicates to facilitate the partial rotation of the sample for regular surveys such as a quarterly LFS.

## 4. Stratification of the Master Sampling Frame in Abu Dhabi Emirate

In order to increase the efficiency of the sample design for the household surveys based on the master sample, it is important to divide the sampling frame of EAs into strata that are as homogeneous as possible.

The nature of the stratification depends on the domains of analysis as well as the most important characteristics to be measured for the different household surveys. For example, the strata should be consistent with the geographic disaggregation to be used in the survey output tables. It is also desirable to order the EAs within each stratum by certain criteria that are correlated with key survey variables, in order to provide further implicit stratification when systematic selection is used, like ordering the EAs by districts. The District is a geographical space delineated on the ground as well as on maps, separated by main roads, encompassing their components. Abu Dhabi Emirate consists of 78 district occupied by households.

The first level of stratification corresponds to the geographic domains of analysis defined for most of the household surveys that will be based on the master sample. The geographic domains for most of the household surveys are the three regions of Abu Dhabi Emirate. These are consistently used for output of estimates. The second level of stratification, within each region, is the national versus non-national households, which will also be an essential domain of analysis.

In the case of some household surveys, SCAD also considered an option of having the districts as individual domains. However, only large-scale surveys would be suitable and these are hard to justify, and therefore infrequent. The master sample is designed to cover the general domains for most of the household surveys, so if a particular survey will have smaller sub-national domains such as districts, they can be based on the full sampling frame rather than be limited to the master sample.

#### **4.1 The stratification strategy by Household type in Abu Dhabi Emirate.**

The stratification technique is frequently employed in the sample design. On the basis of information available from the frame, the units are allocated to strata by placing within the same stratum those units which are more- or-less similar with respect to the characteristics being measured. A stratification strategy was implemented to improve representation of national households in the master sample. This strategy considered the percentage of households with a national household head as a stratification variable. EAs were stratified into different groups based on the percentage of national households in each EA.

Four nationality strata were established according to the ratio of national households in the EAs, these strata are the quartiles of the percentage of national household in the EAs. The nationality strata areas follow:

- Stratum1: 0 to 24.9% national households
- Stratum2: 25 to 49.9% national households
- Stratum3: 50 to 74.9% national households
- Stratum4: 75% or more national households

Since a different sampling rate will be used for each of these strata, the explicit stratification of the master sample will be by region and nationality.

Within each region by nationality stratum, the districts and EAs within each district were ordered geographically in a serpentine manner in order to provide additional implicit stratification when the sample is selected systematically with probability proportional to size (PPS). The geographic ordering of the EAs can also consider the relative socio-economic classification of the EAs.

In addition, the EAs within the highly urbanized districts such as Abu Dhabi Island were classified into three socioeconomic groups (low, middle and high) using maps and local knowledge of the areas. Some areas are mixed, and this classification does not have to be exact. However, this procedure could be used for ordering the PSUs in each socioeconomic group together with the geographic ordering for the implicit stratification within the stratum.

#### **4.2 Sample Size and Allocation for Master Sample EAs**

A master sample is a large sample of *PSUs* for countries that have major and continuing integrated survey programs. The large sample is intended to provide enough “banked” sample cases to support multiple surveys over several years without having to interview the same respondents repeatedly.

The sample size for individual household surveys based on the master sample is determined by the accuracy required for the survey estimates for each domain, as well as by the resource and operational constraints. The accuracy of the survey results depends on both the sampling error, which can be measured through variance estimation, and the non-sampling error, which can only partially be measured through re-interview or validation studies. The sampling error is inversely proportional to the square root of the sample size. On the other hand, the non-sampling error may increase with the sample size, since it is more difficult to control the quality of a larger operation. It is therefore important that the overall sample size for each household survey be manageable for quality and operational control purposes. The sample size for each survey also depends on cost considerations and logistical issues related to the organization of the teams of enumerators and the workload for the data collection.

One of the main determinants of the sample size for each household survey is number and size of the geographic domains that are defined for the survey tabulations and analysis. As indicated previously, the master sample will be designed for household surveys with nationality and regional-level domains. The number of PSUs in the master sample should be sufficiently large to provide sample households for the various household surveys that are being planned.

Although the household survey program is still being developed, the LFS was used as an example of the type of socio-demographic survey that will be based on the master sample. The master sample of PSUs should be larger than the sample size requirements for these particular surveys in order to meet the long-term sampling needs and to allow for future surveys that may require a larger sample size. Most household surveys will be based on a subsample of the master sample PSUs, but some surveys may use the entire master sample PSUs if necessary. For example, some household surveys in the future may cover a small subpopulation group like disabled persons. In

this case, a larger number of sample PSUs will make it possible to conduct a listing to identify households with disabled persons, to ensure a sufficient sample size for the target population. Based on these criteria, a master sample of 560 EAs should be effective for the future household surveys in Abu Dhabi Emirate. This represents nearly 30% of all EAs used in Census 2011, as in the following table:

**Table 2.** Allocation of Master Sample EAs and Replicates by Region

Region	Proposed sample EAs	Number of sample EAs per replicate
Abu Dhabi	320	80
Al Ain	160	40
Al Dhafra	80	20
Emirate	560	140

The allocation of the master sample EAs by region is based on the experience from the previous household surveys as well as future requirements. First, a proportional allocation of the master sample EAs was examined, since this is effective for Emirate-level estimates. In studying the sampling errors from the 2007 HIES and the 2008 LFS, it was found that the previous proportional allocation of the sample did not provide a sufficient level of precision for Al Dhafra. The number of sampled EAs was increased for the region of Al Dhafra to ensure that a sufficient level of precision can be obtained for key indicators from the different household surveys for this region.

It is also recommended that the number of master sample PSUs for each region be a multiple of four in order to establish four master sample replicates that are representative at the regional and Emirate levels. The sample for a particular household survey can be based on a subsample of replicates, or a subsample of EAs within all replicates. Table 2 presents the proposed allocation of the master sample PSUs by region and replicate.

### 4.3 The Proposed Allocation Criteria by nationality strata

The theory of stratified sampling considers the properties of the estimates from a stratified sample and the best choice of the sample size  $n_h$  to obtain maximum precision.

The next step is to allocate the sample PSUs of each region by the four nationality strata identified previously. This can be accomplished by applying a weight related to the proportion of the households in each stratum as shown in the table below.

**Table 3.** Distribution of total Households and National Households by Nationality Stratum

Nationality Stratum	Total No. of EAs	% National HHs	% Total HHs
(1) 0 to 24.9%	1315	7.3	69.9
(2) 25 to 49.9%	255	35.8	13.1
(3) 50 to 74.9%	199	61.7	11.1
(4) 75% or more	114	83.7	5.9
Total	1883	21.5	100

The weighted proportional allocation can be expressed as follows:

$$n_h = n \times \frac{W_h \times \frac{N_h}{N}}{\sum_{h=1}^4 \left( W_h \times \frac{N_h}{N} \right)}, \dots \dots \dots (1)$$

Where:

- $n_h$  = number of sample EAs allocated to stratum h.
- $n$  = total number of sample EAs
- $W_h$  = weight assigned to nationality stratum h
- $N_h$  = total number of households in nationality stratum h
- $N$  = total number of households in frame

Naturally, it is assumed that the four nationality strata will not have the same weights. Since nationality stratum 4 has the highest concentration of national households (75% or higher), it would be assigned the highest weight. The next highest weight would be for stratum 3 (with 50 to 74.9% national households), and the lowest weight (for example, 1) would be assigned to stratum 1. At the same time, the ratio of the highest and lowest weight should be limited to about 4 to avoid extreme variability in the final probabilities of selection and corresponding weights.

Based on equation (1) with:  $\frac{N_1}{N} = 0.699$ ,  $\frac{N_2}{N} = 0.131$ ,  $\frac{N_3}{N} = 0.111$ , and  $\frac{N_4}{N} = 0.059$ . For all possible alternative quadruples of  $W_h, h = 1.2.3.4$ , the expected ratio of the national household in the sample ( $P$ ) was

calculated, as in the following equation:

$$p = \sum_{h=1}^4 \frac{W_h V_h R_h}{\sum_{h=1}^L W_h V_h} \dots \dots \dots (2)$$

Where:

$W_h = 1,2,3,4$ , for  $h = 1,2,3,4$  respectively.

$V_h = \frac{N_h}{N}$ ,  $h = 1,2,3,4$

$R_h$  : Percentage of national households in the stratum (h), it is 0.0725, 0.358, 0.617, 0.837, for  $h = 1,2,3,4$  respectively.

The weights identified to the four nationality strata by the quadruple: ( $W_1 = 1$ ,  $W_2 = 2$ ,  $W_3 = 3$ ,  $W_4 = 4$ ), has produced  $p = 35.6\%$  of national households in the master sample, this is considered high compared with the results for the other quadruples. It also meet the goals of the stratification to have a sufficient sample of national households for each region.

## 5. Conclusion

- The design of the master sample shows that the percentage of national households in the enumeration areas is an efficient variable for stratification. The formation of strata based on this variable has achieved the optimality of the sample design through forming strata characterized by homogeneity of the nationality variable.
- The modified allocation criteria has increased the expected percentage of national households in the master sample, from 21.5% in case of proportional allocation, to 35.6% in the modified one.
- As a result, the construction of the master sample design based on the proposed stratification strategy, and the modified allocation criteria submit an optimum sample be able to present the national household with high level of precision.
- The labor force survey LFS2012 was a case study for implementation of the Master sample. The sample of LFS2012, produced desirable percentage of nationals in each region sample, and achieved the expected level of the relative coefficient of variation for the key estimates in each region.

## References

PLAN ABUDHABI 2030, General Secretariat Executive Council, Emirate of Abu Dhabi.  
 Anthony G. Turner, Sampling frame and master samples. Expert Group meeting to review the draft handbook and designing of Household sample surveys, Dec2003.  
 United Nations, Designing household survey sample: Practical Guidelines, series F No.98. Dept. of Economic and Social Affairs, New York 2005.  
 Des Raj., (1972) "The design of sample surveys ". McGraw-Hill Book Company.  
 Cochran, W.C, (1977) "Sampling techniques " 3<sup>rd</sup> edition, John Wiley, New York.