METHODOLOGICAL DESIGN OF THE ST. LUCIA SAMPLE FRAME TO ASSIST WITH THE CONDUCT OF SURVEYS OF LIVING CONDITIONS AND OTHER MULITPURPOSE SURVEYS

BACKGROUND:

The Central Statistical Office (CSO) of St. Lucia conducted their Census in mid 2001 and have produced a comprehensive Census 2001 database. A preliminary census report was released soon after the census was completed. This report compiled information primarily from the Census Visitation records and it gives the fullest account available of the total numbers of households, dwellings, buildings, non-institutional population of all the census reports. In addition, a more comprehensive report on the findings of the Census of 2001 was printed in April of 2002.

The CSO has engaged in a number of surveys over the years including the Survey of Living Conditions in 1995, the Household Budget Survey 1998, in addition, the CSO has conducts a continuous quarterly labour force survey. A number of other multipurpose social and economic surveys have also been conducted over the years. The CSO has access to a cadre of trained enumerators from the Census and a pool of persons from which it draws in the conduct of multipurpose surveys it undertakes from time to time.

The CSO has had previously a blueprint for the extraction of a sample from a sample frame whenever the need to draw a sample arose. This master sample frame was based on the 1991 census and was done at the time primarily to allow the conduct of Labour Force Surveys in the Mid-1990's. However, this master sample frame was never fully and not enough attention was paid to its maintenance although it was heavily used by staff members of the St. Lucia Central Statistical Office. There are many reasons for this problem as far as the use of the sample frame is concerned not least of which include, the partial existence of a continuous household survey capability programme, the planning of sample surveys based on need or availability of funding for the conduct of such activities amongst others. Therefore, this new blueprint is urgently needed in the context of this MECOVI project and also for the use of the CSO at times when the extraction of an appropriate sample is required as will undoubtedly occur in the affairs of the CSO from time to time. However, training in its use and maintenance overtime, procedures for drawing a sample of specific size from the master sample frame must accompany its institutionalization within the work programme of the St. Lucia Central Statistical Office.

INFORMATION REQUIREMENTS

In order to proceed with the design of the Master Sample Frame for the CSO of St. Lucia the Consultant paid a visit to the CSO to gather some basic information and to consult with staff of the office on technical issues associated with the construction of the sample. The following pieces of information were collected as the basis for informing the design of the Master Sample for the conduct of sample surveys at the CSO in the future:

1. Information on the total numbers of households, buildings, dwellings, persons, both male and female within every Enumeration District in St. Lucia

2. Census Enumeration District level information on percentage of Managers, Professionals and Sub-professionals in population at the Enumeration District level for the purposes of stratifying the sample.

3. Census Enumeration District level information on percentage of workers in Agriculture in the population at the Enumeration District level for the purposes of stratifying the sample.

4. A map of St. Lucia outlining the extent of every Enumeration District and all adjoining enumeration district to it.

The basic intent behind the design of this master sample was to achieve two purposes. Firstly, to allow samples of various sizes to be pulled from the master sample depending on the need. Obviously to fulfill this criterion, a fairly large number of sub-samples/replicates were used as the building blocks for the master sample. This ensured that all districts were represented, this, allowed a measurable probability to be associated with every household. Secondly, the master sample was designed in such a way to allow a continuous sample survey process to be continued. This ensures that the ground work will be laid to allow the quarterly, biannual and annual monitoring and inter period comparison of population aggregates from sample surveys drawn from it.

METHODOLOGICAL CONSIDERATIONS

In developing the sample design of this Master Sample for St. Lucia a number of broad issues were considered. These issues cover four categories:

A. Sampling Errors associated with Multi-stage sampling

The main advantage in the use of the Multi-Stage sampling approach to the development of the Master Sample frame is to reduce cost associated with the use of Simple Random Sampling which would if executed perfectly result in the selection of households far apart from each other in a

random fashion. Multi-Stage sampling employed in the development of the St. Lucia design has the advantage of :

1. Allowing fewer EDs to be selected in the construction of the sample to be used when compared to simple random sampling. This has the benefit of reducing the cost of the sample surveys conducted since fewer areas have to be visited. However, this increases the sample error (or so called cluster or design effect) associated with the survey due to the homogeneity of the households selected within the primary sampling units (Enumeration Districts are the primary sample units in a two stage sample design as is the case with the St. Lucia Master sample frame unless they are joined with adjacent EDs to ensure that every household within EDs have a chance of being selected). As a result of the small size of the islands covered by the MECOVI project, the low levels of transportation cost associated with enumeration activity compared with larger countries and the possibility of improving the precision of sample estimates by increasing the number of EDs sampled, it was decided to design the Master Sample Frame in such a way as to allow larger number of EDs to be selected. This then allowed the numbers of households to be selected within each ED to be reduced substantially when compared to LSMS type samples selected in larger countries, where the cost associated with transport for reaching dispersed households is much greater. In this regard the St. Lucia Master Sampling Frame allows the selection of samples containing approximately five households by systematic random sampling within each selected primary sample units (or EDs).

2. Since the basis for the organization of households in a census is the ED, the ED is used as the primary sampling unit for the organization of the sample.[1] This organization allows the EDs to have a selection probability proportional to its size. The basis for the selection of ED's was the use of a constant sampling fraction of 1/16. The means used to actually select the sample based on this sampling fraction will be elaborated later in this methodological note.

3. In the selection of the households or ultimate sample units, systematic random sampling is used again to reduce the sample error associated with the "design effect" caused from the use of this complex two-stage sampling approach.

B. Geographic Domains, Stratification within the Master Sample Frame

In the St. Lucia sample design two methods of stratification are used. Firstly, to improve the overall precision of the estimates of population proportions obtained from the sample a listing of all enumeration districts (ED) was developed for each district, this list was stratified using either the % of Managers, professionals and sub-professionals for each ED within urban districts or the % of Agriculture workers in the EDs located within the rural districts. The procedure employed will be fully elaborated in the paragraphs that follow. Secondly, the sample was partitioned into districts to ensure that these important subgroups in the population were all properly represented.

[1] In a small number of cases PSUs are formed by a combination of very small EDs or a combination of a very small and a large ED, EDs described here as being small contain less than 30 households within them.

For the SLC (surveys of living conditions) it was essential that it be possible to obtain reliable estimates of population proportions for all of the districts for policy, political and other reasons.

ELEMENTS OF THE ST. LUCIA MASTER SAMPLE FRAME

The design of the grand sample for St. Lucia draws on the sample design of the Trinidad and Tobago CSSP (Continuous Sample Survey of Population) which was first operationalized in January of 1987 and remains in existence today while having been updated with information from the Trinidad and Tobago Censuses of 1990 and 2000.

The basic design contains the following features, a Master sample frame consisting of nine (9) replicates/sub-samples (clusters of households) are selected with a probability of 1:16. The choice of this large number of replicates was deliberate. The use of nine replicates allows the construction of samples of different sizes from the Master sample frame. This is ideal for SLC or LSMS type sample survey activities which tend to be a one off activity repeated from time to time based on need and available financial resources. In addition, the use of nine (9) replicates allows for the rotation of sub-samples/replicates and the collection of data on a sample of households on a continuous basis, either quarterly, biannually or annually. Population estimates from continuous samples done in St. Lucia can therefore be generated after the completion of each round of the survey.

The Multi-Stage sampling procedure developed for the St. Lucia MS (Master Sample) Frame

As has been described previously the two stage process of sample selection in the ST. LUCIA MS entails the selection of the PSUs within the districts. This is followed by the systematic selection of the cluster of households or USU (Ultimate Sampling Units) within the selected PSUs. The two stages in the design is elaborated as follows:

a. In the first stage, a sampling frame is constructed consisting of all of the enumeration districts from the census of 2001. The size of each enumeration district is measured in units of clusters of households. In the case of the ST. LUCIA MS, approximately five households were allocated per cluster. The clusters which are allocated to the EDs all have an equal probability of selection within the specified geographic domain in which they are allocated. In addition, the number of clusters allocated to an ED is a measure of the size of the ED. Clusters, therefore ensure the selection of EDs or Primary Sampling Units with probability proportional to the size of the ED. The ST. LUCIA MS frame consists of nine sub-samples / replicates, with each replicate selected with a probability of (1 / (16 * 9)) or 1 / 144.

b. In the second stage a non-compact cluster of households is selected within the selected PSU[2] using systematic random sampling. There are three elements to the selection of this non-compact cluster. Firstly, there is the sample interval, which is a measure of the size of the ED in terms of the total number of households it contains. The larger the ED or PSU the larger will be the sample interval assigned and consequently the larger will be the number of clusters assigned to the ED. This approach ensures that the total number of households selected in any selected ED is approximately the same. In the case of the "Castries" in the ST. LUCIA MS frame the approximate number is five (5). Secondly, the random start is determined by use of a random number generator. With a Microsoft EXCEL spreadsheet the formulae takes the following form, =ROUND(RAND()*E1,0)+1, where E1 is the cell containing the sample interval (or total number of clusters assigned) RAND() is the function which generates the random number. The round() function is used to round the result to the nearest whole number. The third element of choosing the non compact cluster is a combination of the above. A random number (r) is choosen between 1 and the sample interval value, I, inclusive, then to this number is added the sample interval for the full list of households within the primary sample unit. Thus, the list of selected households would be r, r + I, r + 2I, r + 3I, r + 4I,……, r + (n – 1)I, where n is the cluster size assigned to the district, in the case of Castries n is five.

[2] The term PSU is generally used interchangeably with the ED and these terms are generally the same. However, where large EDs must be broken down into segments to improve the precision of sample estimates, the segments are the PSUs and not the EDs in those instances. Similarly, small EDs may be joined to adjacent EDs to form larger PSUs.

A. Size of the Sample

As has been explained before the decision to use a sampling fraction of 1 : 16 and to assign nine replicates to each District (the geographic domain) was based on the need to take advantage of the small size of the countries covered by this MECOVI project. This was done by increasing the "spread" of the sample across EDs and as a result improving the precision of the estimates which can be obtained from it. In addition, attention was paid to ensuring that were the CSO of ST. LUCIA to consider developing further its Integrated Household Survey Programme, the ground work would have been laid through this Master Sample Frame design for periodic, ad hoc or continuous sample surveys. The achievement of this objective has already been demonstrated through the use of this Sample Frame in the conduct of St. Lucia's continuous Labour Force Survey.

Therefore for any one sub-sample given that there are nine, the sampling fraction is 1 / 16 by 1 / 9 or 1 / 144. If a periodic, ad hoc or quarterly survey included the use of three replicates then the sampling fraction for these three replicates would be 3 / 144 or 1 /16 by 3 / 9. In both cases the resultant sampling fraction is the product of the sampling probability for the Master Sampling frame and the probability of selection of a specific number of replicates.

B. Master Sample Domains of Study and Stratification

1. Domains of Study:

The Master Sample frame was subdivided into eleven areas for the purpose of the provision of estimates from samples selected from this frame. The following list of the ten domains or sub-populations is based on the Districts which formed the basis for the collection of information on the population in the 2001 Census.

TABLE 1 : Domains of Study and Stratification criteria

DOMAIN OF STUDY	No of PSUs	METHOD OF STRATIFICATION
Castries Urban	42	Managers, Professional, technical and related Administrative
Castries Rural	112	Managers, Professional, technical and related Administrative
Anse-La-Raye/Canaries	24	Workers in Agriculture
Soufriere	22	Workers in Agriculture
Choiseul	20	Workers in Agriculture
Laborie	18	Workers in Agriculture
Vieux-Fort	37	Managers, Professional, technical and related Administrative
Micoud	42	Workers in Agriculture
Dennery	29	Workers in Agriculture
Gros-Islet	55	Managers, Professional, technical and related Administrative

The total number of PSUs in the ST. LUCIA MS is 401, a breakdown of the number of PSUs by District is shown in the table above. The average size of the PSUs was 118 approximately with a standard deviation of approximately 47. This configuration does not in the near term present a major problem for sample implementation, since the EDs/PSUs size does not exceed 100 by too great an extent, in addition, while consideration must be given to splitting EDs which have grown in size to over 200, there are not as exist in the case of St. Vincent and the Grenadines a significant number of excessively large EDs. Continuous maintenance of this situation is required and can be done by splitting all EDs over 200 in size into smaller ones of approximate size 100. The main objective of controlling the size of the PSUs, is to reduce variability and thereby improve the precision of estimates from the sample. The more equal the sizes of the PSUs the more likely the variance of characteristics between PSUs will be minimized and inversely the precision of the samples derived from the estimates from the Master Sample Frame increased[3].

2. Stratification

As shown in the table above each of the domains of study was stratified according to specific criteria. In the more urban domains the criteria used was the percentage of Managers, professional, sub-professionals in the population. The PSUs or EDs were therefore arranged in descending order of the proportion of this group in the population of the ED. In the rural domains the PSUs were arranged in descending order of the proportion of agriculture workers in the population of the ED. In the case of Canaries and Anse-la-Raye, the sizes of the populations in these domains mandated a joining of the two to allow for the creation of a large enough domain for reporting purposes.

CONSTRUCTION OF THE MASTER SAMPLE FRAME

A: Sectors and the assignment of Clusters to Domains

In order to construct the master sample frame a fixed number of ultimate sample units or cluster size has to be associated with each domain of study. In order to pre-determine the cluster size two conditions were satisfied. In order to allow the sample to generate quarterly results if desired, the first condition was to ensure that three replicates/sub-samples could be enumerated per month within a specific quarter. Secondly, the total number of PSUs selected for enumeration per month or per quarter had to be exactly the same on every occasion when a monthly or quarterly selection is made.

To achieve these conditions, the sample frame was designed to ensure that the total number of clusters assigned per domain was an exact multiple of (3 * 9 * 16 = 432). Three (3) represents three months in a quarter, nine (9) the number of replicates/sub-samples in the master sample frame and 16 the sample fraction associated with the Master Sample Frame. Domains were subdivided into sectors derived from an examination of the percentage distribution of the population. In the table which follows the process by which sectors were assigned to domains is shown.

[3] (See Kish, L. (1965), pp. 184 – 189, 218 - 219

	No of PSUs	No of Hholds	% of Total	Population	% of Total Population	Avg Hhold Size	Sectors Assigned	Cumulative No of Sectors	No of Clusters	Average Cluster size of household
Castries Urban	42	4507	9.53%	14800	9.4%	3.28	2	864	864	5.21644
Castries Rural	112	14943	31.61%	49545	31.4%	3.32	7	3024	3888	4.94147
Anse-La-Raye/Canaries	24	2401	5.08%	7848	5.0%	3.27	1	432	4320	5.55787
Soufriere	22	2252	4.76%	7656	4.9%	3.40	1	432	4752	5.21296
Choiseul	20	1728	3.66%	6128	3.9%	3.55	1	432	5184	4.00000
Laborie	18	2012	4.26%	7363	4.7%	3.66	1	432	5616	4.65741
Vieux-Fort	37	4144	8.77%	14754	9.4%	3.56	2	864	6480	4.79630
Micoud	42	4588	9.70%	16041	10.2%	3.50	2	864	7344	5.31019
Dennery	29	3753	7.94%	12767	8.1%	3.40	2	864	8208	4.34375
Gros-Islet	55	6947	14.69%	20872	13.2%	3.00	4	1728	9936	4.02025
	401	47275	100.00%	157774		3.34	23	9936		4.76
							23			
Sampling Fraction for the St. Lucia Labour Force Survey					1 / 16			48		144
First Stage Sampling Units are selected with a probability of:					(1/(16*9))			1		1
The Sampling Fraction for each quarter of the survey is:					(1/16 * 3/9)	=	(1 / 48)	0.0208		
Sampling Fraction for each replicate					(1/16 * 1/9)	=	(1 / 144)	0.006944444		
For three months in a given quarter for three replicates sampled					207					
Number of Eds visited per month					69					

Table 2 shows the procedure used to arrive at the 23 sectors allocated across ten domains of study based on the percentage distribution of the population across the domains. As an example, Castries Urban with 9.4% of the population was assigned two sectors each of which was further subdivided into nine replicates or sub-samples. These two sectors represented a total of 2 * 432 clusters assigned to the Castries Urban domain. Since the total number of households in Castries Urban 4,507, the average cluster size for this domain was calculated to be 5.2 households. This average cluster size was then used to determine the number of clusters assigned to each ED or PSU which is a measure of the size of the PSUs and a means of ensuring that PSUs are selected with probability proportional to their size.

B: Selecting the nine replicates in the Master Sample Frame

Since the number of clusters assigned to any domain had to be an exact multiple of 432 (3 * 9 * 16). The number of clusters was allocated to each PSU by dividing the total number of households assigned to PSUs by the average cluster size assigned to the domain. Once this has been done and the total number of clusters assigned to each PSU is established, a random number between 1 and the inverse of the sample fraction sixteen (16) inclusive is assigned, subsequent replicates are choosen by adding 16 to this random number systematically. To illustrate the selection of the first nine replicates for Castries Urban, the first replicate is selected based on a random number 2, the second value 18 (i.e. 2 + 16) lies below or is equal to the cumulative total number of clusters for PSU/ED 00801 of 34 as does the third value 34 (i.e. 2 + 16 + 16). Consequently, sub-samples a is assigned to PSU/ED 00600 and sub-samples d and g are assigned to the second listed PSU/ED 00801. A similar procedure is used to select the remaining clusters of the nine replicates as shown in the table which follows.

Within any allocation of one sector, this selection process must be carried out three (3) times e.g. 3 (16 * 9) = 432. The selection of the first sampling unit for each of the respective nine (9) replicates listed in chronological order e.g. 1001, 2001, ... 9001 constitute a segment, that is, one third (1/3) of a sector. The selection process shown in Table 3 is repeated throughout the domain until the total number of clusters (sampling units) assigned to the domain of study is exhausted. In the supporting spreadsheet contained on the CD ROM enclosed with this documentation an example of the use of the sample frame to draw a 2.08% or (3/144) sample using replicate g, h, i is shown. Similar 2.08% samples can be selected for each of the seven replicates which make up the frame.

TABLE 3 - EXAMPLE OF REPLICATE/SUB-SAMPLE SELECTION FOR THE CITY OF CASTRIES

Castries City

	E.D. NO.	% employed as managers, professionals, sub-professionals	% employed as workers in Agriculture	No. Of Hholds	No. Of Clusters Assigned	Cummulative Total Of Clusters	Replicate Number Random Start=2	Alpha Sample Selection
1	00600	60.00	2.22	33	6	6	1001	a
1	00801	49.12	0.58	143	27	34	2001	d
1	00802	46.04	0.72	102	20	53	3001	g
							4001	h
1	00502	38.35	6.77	109	21	74	5001	b
1	03302	37.50	0.78	137	26	100	6001	e
							7001	f
1	00901	32.93	0.00	115	22	122	8001	i
1	01201	32.05	1.28	70	13	136	9001	c
1	01406	31.29	1.36	147	28	164	1002	a
							2002	d
1	01703	30.86	1.23	69	13	177		
1	01001	30.00	0.00	79	15	192	3002	g
1	00903	29.35	1.09	99	19	211	4002	h
							5002	b
1	00200 + 00300	28.20	0.00	65	12	224		
1	03203 + 03301	28.16	1.15	211	40	264	6002	e
							7002	f
							8002	i
1	01800	27.91	1.55	100	19	284	9002	c
1	01002	27.42	0.81	108	21	304	1003	a

Theoretical aspects of sample selection with probability proportional to size

Given that the procedure for selecting sampling units involves the selection of one cluster of households (ultimate sample units or usus) per psu (ED). The samples have been designed in such a way as to ensure that each domain was allocated sample units, which were an exact multiple of either a subset of households within a domain or the smallest Domain. Ultimate sampling units were then selected with probability proportional to size of the PSU. The overall objective was to obtain a uniform sampling fraction across all domains for each of the countries. Therefore for any domain, the selection of the cluster reduces to the form shown in the Equation which follows i.e. bh/Fh i.e. 1/F since bh (number of clusters to be selected) was constrained to one (1) cluster per PSU. The selection probability for each PSU can be expressed as follows:

$$\frac{Mos\alpha_h}{Fb_h} \times \frac{b_h}{Mos\alpha_h} = \frac{b_h}{Fb_h} = \frac{1}{F} = f$$

where Mos α_h is the original measure of size prior to conversion to clusters (i.e. number of households) in the α^{th} ED (psu) of the h^{th} domain; Fb_h is the zone size, i.e. the product of the inverse of the sampling fraction (f), and b_h is the average cluster size based on allocation of sampling units per domain in the h^{th} domain; f is the overall sampling fraction for replicate sampling in the design of the sample for a survey of living conditions; and, F is the inverse of the sampling fraction of the survey.

Weighting and Non Response Rates

The following table summarizes the outcome of the 1999 St. Kitts Survey of Living Conditions Survey and the method used to re-weight the sample based on non-responses obtained in the process of conducting the survey. This example is applicable since a similar strategy in the design of the sample is employed here as was employed in the conduct of the SLC in St. Kitts and Nevis. “Expected” is the total number of questionnaires expected from the conduct of the survey. “Number Obtained” is the number of questionnaires completed by the enumerators conducting the enumeration exercise. “Response Rate” measures is the number of questionnaires obtained over the number of questionnaires expected on a parish by parish basis with a simple average overall for all parishes. “Weighted Response Rate” is the weighted response rate, the number of questionnaires obtained by parish over the total number of questionnaires expected. This rate gives an indication of the percentage of the total sample expected which was completed in each parish. “Expected Adjusted” gives an indication of the number of interviews expected when the distribution of the population from the 1991 census is considered. “Raising factor” is the amount by which the number of questionnaires/responses obtained will have to be increased to achieve the expected number of questionnaires for the parish. The “Expansion Factor” is the inverse of the sampling rate (i.e. 1/0.1, 10% was the expected size of the sample in the SLC conducted in St. Kitts and Nevis) by the raising factor. When the expansion factor is applied to the survey of living conditions data the total population estimates are derived. Thus when the expansion factor is applied to the number of households obtained the total number of households by parish is the result i.e. the last column of the following table.

National Survey Response Rates for St. Kitts Living Conditions Survey 2000

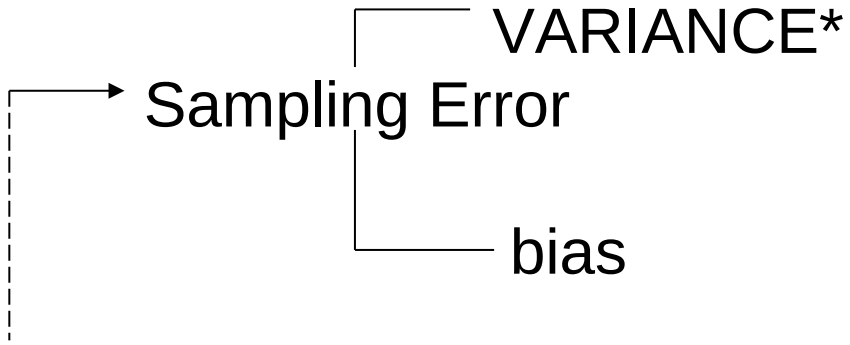
Parish	Expected	Number Obtained	Response Rate By Parish	Weighted Response Rate	Expected Adjusted	Raising Factor	Expansion Factor	Number of Households by District
St. George Basseterre East	215	206	96%	26%	245	1.19	11.89	2450
St. George Basseterre West	138	113	82%	14%	157	1.39	13.89	1570
Trinity	31	34	110%	4%	35	1.03	10.29	350
St. Thomas	52	46	88%	6%	59	1.28	12.83	590
St. Anne	64	59	92%	7%	73	1.24	12.37	730
St. Paul	40	37	93%	5%	46	1.24	12.43	460
St. John	77	97	126%	12%	88	0.91	9.07	880
Christ Church	45	44	98%	5%	51	1.16	11.59	510
St. Mary	99	78	79%	10%	113	1.45	14.49	1130
St. Peter	40	64	160%	8%	46	0.72	7.19	460
	801	778	102%	97%	913			9130

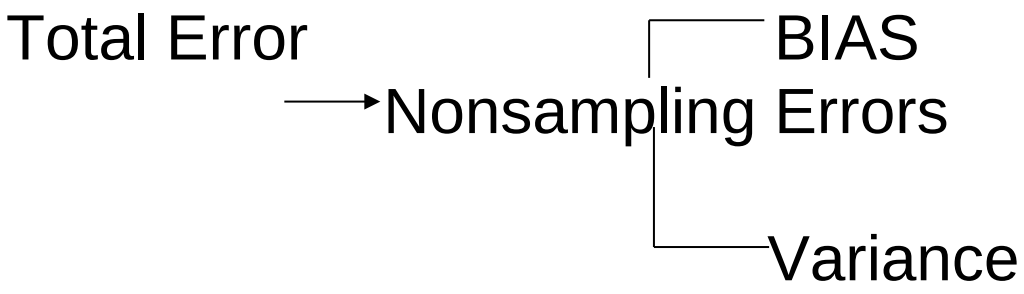
Source: Statistical Department

The difference between the number of questionnaires obtained and the number of questionnaires expected is a combination of refusals, no contacts with the selected households and growth in the population from the Census of 1991 to 1999 when the SLC was completed. Therefore, 97% of the expected number of interviews was completed overall.

Sampling Errors

The sum of all errors affecting an estimate from a sample is known as the 'total error'. The total error is represented mathematically by the mean square error (MSE); its value is usually unknown in practice. This MSE has two components: sampling errors and non-sampling errors (NSE's). In actuality, there is a double dichotomy:





*the main component of sampling error is variance

The sampling error is constituted mainly by variable errors called variance. The variance is the average deviation of sample estimates from the average of all possible estimates under the same sample design. The variance indicates the precision (reliability) of the estimates which is represented by the standard error of the estimate equal to the square root of the variance. The variance is lower the larger the size of the sample and the more efficient its design. The standard error of an estimator Φ is given by:

$$s(\Phi) = \sqrt{\text{var}(\Phi)}$$

For the poor, the non-poor, the headcount this standard error is used to develop confidence intervals to see the range of possibilities for the true value of these poverty indicators existent within an SLC;

$$\Phi \pm 1.96s(\Phi)$$

coefficients of variation (CV) – the relative standard errors—which allows the evaluation of the precision in relative terms and compare precision levels for estimates of different poverty indicators in different populations;

$$CV(\Phi) = \frac{s(\Phi)}{\Phi}$$

The design effect (DEFF) is used as a means of comparing the efficiency of the systematic stratified random sample used to that of simple random sampling design.

$$DEFF = \frac{\text{var}_{ssrs}(\Phi)}{\text{var}_{srs}(\Phi)}$$

Recommendations for Quality Sample Frame Maintenance

Periodic re-listing of enumeration districts

For the sample to serve as a good base for deriving population parameter estimates it must be maintained. One important aspect of the maintenance that must be done is the listing of enumeration districts to ensure that the most recent total numbers of households is inserted in the sample frame. This ensures that the most accurate sample is taken after provisions are made for adjusting replicates and selection procedures based on the structural changes which result from the restructuring of the sample based on updated information derived from the listing exercise. To be most efficient re-listing of households are generally concentrated in the enumeration district which have experienced the most change since the last census upon which the sample frame is derived. Determining the extent of change within an ED depends on a value judgment which is based on physical observation. Re-listing is a simple exercise which involves use and update of an enumeration district map. Under ideal circumstances, a satellite image of the Enumeration district should first be obtained. Then using Geographic Information System software all new buildings are added to the topographic representation of the Enumeration Districts using “heads-up” digitizing. Enumerators when preparing for sample surveys which are conducted from time to time by the Statistical Office will visit the enumeration district and list from start to end based on the route specified by the enumeration district map all buildings, households, addresses, names of household heads and numbers of persons living in the household, both male and female for the Enumeration District. The total numbers re-listed can then be used to re-structure the sample. Please note that as much as possible the listing exercise should be separated from the conduct of sample surveys due to the enumerator bias that can be introduced if the selection of households to be enumerated is done at the same time as the listing exercise.

Splitting of Enumeration Districts

In Grenada and particularly in St. Vincent and the Grenadines this activity needs to be done as a matter of urgency. The reason for splitting an enumeration district has been stated before in this paper, however, this bears repetition. EDs of more or less similar sizes help improve the ease of sample implementation, by reducing the sample interval within EDs and therefore the transportation cost associated with enumeration of households within selected EDs. In addition, the precision of the estimators produced from samples with EDs of uniform sizes obtained from the Master Sample Frame is greater than that obtained from EDs of vastly varying sizes. Therefore every effort should be made to ensure that EDs do not grow in the case of the OECS countries to sizes in excess of 150 and are in most instances best maintained to a size of 100 households. In order to achieve this each of the large EDs must be examined and important landmarks or demarcations identified with a view to splitting the EDs into manageable sizes. When this is done an ED of size 250 which was assigned a sample interval of 25 in the self-weighting sample described above, when split into two PSUs or EDs, for example of sizes 115 and 135, will have sample intervals of 115/250 * 25 = 12 and 135/250 * 25 = 14 respectively. In the case of countries with many large EDs an effort therefore needs to be made to conduct these procedures and if not create EDs immediately but at least segment them to reduce the cost associated with sampling these Enumeration District.

Improving the Technology of Enumeration District Maintenance

Geographic Information Systems (GIS) Technology is a proven and established way of maintaining ED maps. They have eliminated many problems involved in area/panel sampling, specifically, a well designed GIS data model eliminates the possibility of overlapping enumeration districts. GIS also ensures that EDs are maintained overtime in such a way which ensures that as much information as can be made available can be brought to bear on the process of ED maintenance. Digital maps from urban planning agencies, satellite photography and survey base maps which are periodically updated, enumerator map updates, field inspections by staff of the mapping unit can all be incorporated into a GIS to enhance the mapping and by extension listing and maintenance of a good sample frame.

Summary

The author of this paper hopes this work has created a framework for the conduct of sample surveys in the respective islands which it covers. It should serve as the base or foundation from which a well designed sample survey can be derived on a systematic basis over the quarters of a year, annually or an ad hoc basis in the countries of the OECS to which it applies.

REFERENCES

COCHRAN, W. G. (1977) Sampling Techniques, New York: John Wiley.

HANSEN, M. H., W. N. HURWITZ, and W. G. MADOW (1953) Sample Survey Methods and Theory, Vols. 1 and 2. New York: John Wiley

KALTON, G. (1977) "Practical methods for estimating survey sampling errors." Bulletin of the International Statistical Institute: 47, 3: 495-514

KALTON, G. (1983) "Introduction to Survey Sampling." Sage University Papers, 35, Series: Quantitative Applications in the Social Sciences

KISH, L. (1965) Survey Sampling. New York: John Wiley.

MOSER, C. A. and G. KALTON (1971) Survey Methods in Social Investigation. London: Heinemann

TRINIDAD AND TOBAGO, Central Statistical Office (1987), "Revised Design of the Continuous Sample Survey of Population (C.S.S.P.) Methodology Report, 1987"