

PN ABY-709
90772

**SECRETARIA DE RECURSOS NATURALES
UNIDAD DE PLANEAMIENTO, EVALUACIÓN Y GESTIÓN (UPEG)**

INSTITUTO HONDUREÑO DEL CAFE

**EVALUATION OF THE IHCAFE STATISTICAL SYSTEM TO FORECAST
COFFEE PRODUCTION IN HONDURAS**

Prepared by: Clarence Dunkerley

*Economic Policy Analysis Project (No. 522-0325)
USAID Contract Number: 522-0325-C-00-3298-00
June 1996*

PRODEPAH is a project of the United States Agency for International Development conducted in conjunction with the Government of Honduras and carried out by Chemonics International Inc. in collaboration with Sigma One Corporation

TABLE OF CONTENT

LIST OF TABLES		iii
ACRONYMS		iv
SECTION I	INTRODUCTION	1
SECTION II	FINDINGS	4
SECTION III	RECOMMENDATIONS	6
SECTION IV	CONCLUSION	13

LIST OF TABLES

Table 1.1	Coffe: Production Forecast, Production and “Registros de Compra”, Honduras, 1987/88 to 1995/96.	2
Table 2.1	Stratification of Coffe Producers Registry, Honduras, 1996.	5
Table 3.1	Sample Size of Coffe Producers at the National Level, Honduras, 1996.	6
Table 3.2	Proportional Distribution of “n”, by Number of Producers by Region, CV=10% and CV=5%, Honduras, 1996.	7
Table 3.3	Sample Size To Make Estimates of Coffe Production at the Regional Level with a CV=20%, Honduras, 1996.	8
Table 3.4	Sample Size To Make Estimates of Coffe Production at the Regional Level with a CV:10%, Honduras, 1996.	8
Table 3.5	Sample Size To Make Estimates of Coffe Production at the Departamental Level with a CV=10%, Honduras, 1996.	9
Table 3.6	Sample Size To Make Estimates of Coffe Production at the Departamental Level with a CV=20%, Honduras, 1996.	10

ACRONYMS

ASE	Area Sampling Frane
CV	Coefficient of Variation
IHCAFE	Instituto Hondureño del Café
USAID	US Agency for International Development

SECTION I INTRODUCTION

In 1988 IHCAFE began development of a statistical system, based on probability sampling techniques, to forecast expected coffee production for the coffee year before harvest begins. This system was based upon a multiple frame sampling technique that used a list frame of approximately 10,000 small coffee producers (beneficiaries) who agreed to produce and improve their productivity under the guidance of the IHCAFE small farmer improvement project. In addition an Area Sampling Frame (ASF) was constructed to estimate for the rest of the coffee producers (non-list) in the sector. The ASF insured complete coverage of the Honduran coffee sector.

The samples were selected within each of the nine IHCAFE regions from both the List Frame and the Area Sampling Frame. The ASF required the interviewers to first locate boundaries drawn on topographic maps (cartas) of one to two square kilometer sample areas (segments) and then to interview all coffee producers that lived within those boundaries. All producers living within those boundaries were identified and then they were matched against the 10,000 names on the list frame. Only those that did not match (non-beneficiaries) -- e.g. they were not listed as a beneficiary -- were counted for the ASF sample. Thus, all national estimates were the sum of the two mutually exclusive domains. The estimator was $Y_{total} = Y_{list} + Y_{area\ frame}$.

This survey process required rigid control of the data collection process. This means the interviewers (IHCAFE Extension Agents) were made familiar with the questionnaires, interview methodology, map reading techniques, and procedures to positively identify the segment boundaries. Prior to each survey the Statistical Section of IHCAFE spent a day in each regional office training the interviewers and giving them the required materials to complete their assignment. During the survey the central office supervisors visited each area to assist and insure that the interviewers were following the correct procedures. Basically three people (two Honduran technicians and one USAID advisor) organized and completed this work.

The first survey was carried out in August of 1988. Subsequent surveys were done each year to forecast the coming harvest. In addition, special surveys were done to investigate the social and economic structure of the coffee sector.

The USAID statistical advisor to this project was reassigned in May 1991 and the project remained under the management of the IHCAFE Statistics Section. They continued the data series by completing surveys in 1991, 1992, 1993, 1994, and 1995. The original technicians trained for this activity left the Section for other employment in 1992, and since then, training of the interviewers and supervision of the data collection has been lacking. This needs to be addressed in the near future.

In August, 1992 IHCAFE began a program to register all producers of coffee in Honduras in order to pay the producers a bonus from the accumulated coffee tax funds. Approximately 80,000 producers were registered in 1992/93. This list is updated annually as coffee producers sell their production which is registered in the "registry de compras". This same "registry de compras" has served as the primary source of sales data to evaluate the production forecast from the survey data. Table 1.1 illustrates the survey data compared to the final sales data in quintales oro as registered by IHCAFE in its "registry de compras". The forecast of production column represents what the producers anticipate they will harvest in the forthcoming harvest season, and the production obtained column is the production they actually harvested during the previous harvest season.

Table 1.1 Coffee: Production Forecast, Production and "Registros de Compra", Honduras, 1987/88 to 1995/96

Crop year	Production Forecast	Production	Registros de compras
	(000)	(000)	(000)
1987/88		1934	1923
1988/89	2254	2170	2030
1989/90	2515	2550	2453
1990/91	2150	2303	2045
1991/92	3078	3167	2794
1992/93	2575	2450	2487
1993/94	3229	2650	2414
1994/95	3408	2938	2728
1995/96	3030*		

*forecast

Source: IHCAFE.

This table demonstrates that the survey data are reliable predictors of the final registered sales. One would expect the survey estimates of production to always exceed the "registros de compras" data since the survey data are estimates of total biological production. Biological production includes that production the producer retains for home consumption, seed, storage, losses, non-registered sales, and registered sales.

The Beneficiary List served as the initial list sample frame. The small farmer project has been discontinued since 1991, thus, its utility as a sampling frame becomes more problematical. It can not be updated and respondents in the area frame segments are more likely to say they were not beneficiaries of the project making it more likely to over-estimate the survey results.

The Statistical Section of IHCAFE has been using personnel computers (PCS) to manage and summarize the survey data. The Project began with the LOTUS 123 spreadsheet package with programs and procedures developed to do data input, data edits, data summarization, and generation of output tables. Beginning with the 1995/96 survey everything was changed to FOXPRO, a new database package that is more efficient and flexible.

The Statistical Section was established in 1988 with the assignment of three people to the project. Two of the three were experienced in sample frame construction and survey management. The third person came into the project without experience but was trained by the others in the section and proved to be very capable. By 1992 all three experienced people had left IHCAFE for other work, leaving IHCAFE without experienced personnel to manage the statistics program.

SECTION II FINDINGS

This consultant, who was the original advisor to the IHCAFE statistics project, was well received by the present IHCAFE staff, many of whom I had worked with previously. Having intimate knowledge of the development and management of the statistical program I was able to quickly review the current procedures and the historic data series.

It was readily apparent that changes would have to be made to the list sample frame if this methodology is to continue. The list frame should be shifted from the list of beneficiaries, about 10,000 names, to the "registros de compras", which has 90,000 names thus giving more coverage of the coffee producing sector. The "registry" is also updated annually.

I had a difficult time trying to review the survey data for two years - 1993/94 and 1994/95. There were problems with the data and summary diskettes which prevented us from looking at the internal summaries. I was told that no interviewer training was done these two years which could have contributed to a decrease in the reliability of the information released from this survey. There appeared to be a greater than normal increase in land area from which coffee was harvested when comparing 1992/93 to the following two years. At this point it is difficult to know or understand the reasons for this change in the level of acreage dedicated to coffee.

The 1994/95 survey was conducted after some (minimum) training of the interviewers in the regional offices. The results of this survey appear to be more consistent with the earlier data series.

For the first time since 1992 a Quality Control survey was done following the most recent survey. In October, 1995 nineteen Area Frame segments were visited to verify the previous month's work of various interviewers. Seven of the nineteen segments were totally incorrect - either having not been visited or adding people to the segment that did not live in the area. Four segments were partially correct having some of the same respondents living in the segment. Eight were done correctly. The major problem is that the enumerators are not well trained in reading, understanding and using cartographic materials and have great difficulty in locating the segment boundaries. This is extremely difficult work in exceedingly mountainous terrain. The new updated maps (cartas) will partially solve this problem.

I reviewed the "registry nacional de productores de cafe" (registry) and think that it will be an excellent sampling frame to complement the Area Sampling frame. It is current, has auxiliary data associated with each record, has two identity fields (Identidad y Clave de Productor) that greatly assists with matching names, and is easily accessed and manipulated because it is in the FOXPRO database system.

The survey data processing has also been shifted to FOXPRO. A data input program has been written that is easy to use for the transcribers and the report generating programs greatly accelerate the processing and preparation of the final report. This FOXPRO system being used in IHCAFE has been well thought-out and prepared here in IHCAFE. The capability already exists to sort and stratify the list, select samples from each sorted stratum, calculate expansion factors, match the Area Frame sample producers with the List Frame to determine the duplication between the two frames. Based on my observations of the "Resumen Estadístico" of the "registry", I decided to stratify the 90,000 producers into five strata based on their production.

Table 2.1 Stratification of Coffe Producers Registry, Honduras, 1996.

Stratum	Limits of the stratum qq	Number of Producers	qq of Production	% of Total Production
I	0 < 15	41,685	266,367	6.6
II	15 < 50	30,379	820,233	20.4
III	50 < 300	17,365	1,771,230	44.1
IV	300 < 1000	1,701	768,852	19.1
V	1000 +	195	395,792	9.8
Total		91,325	4,022,474	

It can be seen that 46% (41,685) of the producers produce only 6.6% of the total production. This is a good place to introduce the fact that IHCAFE has never defined a coffee producer. The present procedure includes everyone that produces or sells even less than one quintal or has a garden with a few coffee bushes that are used for home consumption. It can also be seen that only 2% of the producers are responsible for almost 30% of the Country's production. This fact greatly influences the sample design. The stratum were designed to put producers into homogeneous groups to minimize the within stratum variance.

SECTION III
RECOMMENDATIONS

1. I recommend that IHCAFE change the source of the List Frame from the Beneficiary list to the “registry nacional de productores” list. Table 3.1 shows the division of the list into 5 strata and the number of producers on the list within each stratum, as well as average production and population variance per stratum. I show two columns for sample size “n”. The first gives the necessary sample size if you wish to estimate the “true population” mean within plus or minus 10%, given a 1 in 20 chance of selecting a sample that would give a result outside these limits. The second column shows the necessary sample size if the required precision was changed to plus or minus 5%. Note that the sample size is almost 4 times greater for the higher (5%) precision sample.

For example the following list sampling scenarios can be shown:

- | | | | |
|----|------------------------------------|----|---|
| 1. | National estimates only | 1. | A simple random sample from the entire list. |
| 2. | Regional and National estimates | 2. | Sort list into regions and then select the samples from each region. |
| 3. | Department and National estimates. | 3. | Sort the list into departments and select Samples from each department. |

Table 3.1 Sample Size of Coffe Producers at the National Level, Honduras, 1996.

Estratum	Number of producers	qq Mean	Variance	sample size	
				n cv=10%	n cv = 5%
I	41,685	6.4	17.4	161	637
II	30,379	27.0	87.2	46	183
III	17,365	102.0	2,839.7	104	410
IV	1,701	452.0	24,961.9	46	169
V	195	2,029.7	2,938,768.	114	166
total	91,325			471	1,564

Table 3.2 shows the allocation of the two samples to the 9 IHCAFE regions. This allocation is based on the proportion of producers within each stratum in the region.

Table 3.2 Proportional Distribution of “n”, by Number of Coffe Producers by Region, CV=10% and CV=5%, Honduras, 1996.

Stratum	Region									Total
	1	2	3	4	5	6	7	8	9	
cv=10%										
I	17	32	14	19	16	18	20	12	12	160
II	4	10	4	7	5	4	6	4	2	46
III	11	23	6	18	12	8	12	11	4	105
IV	6	10	2	9	5	5	3	5	2	47
V	16	18	5	22	12	16	5	17	2	113
Total	54	93	31	75	50	51	46	49	22	471
cv=5%										
I	67	127	56	74	64	72	78	49	49	636
II	18	39	14	26	20	15	24	17	9	182
III	44	90	22	71	46	32	47	42	15	409
IV	23	36	5	33	19	19	10	20	5	170
V	24	26	8	31	18	24	7	25	3	166
Total	176	318	105	235	167	162	166	153	81	1563

Tables 3.3 and 3.4 show the sample sizes and allocation to strata within regions in order to estimate at the regional level with a level of precision of 10 or 20%. This would produce national level estimates with coefficients of variation (CV's) between 5 and 10% of the true population means. If one wanted regional estimates with precision (CV) near 10% a sample of approximately 3200 producers from the list frame would be required. This would probably put too much of a demand on the available resources in IHCAFE.

Table 3.3 Sample Size to Make Estimates of Coffe Production at the Regional Level with a CV=20%, Honduras, 1996.

Stratum	Region									Total
	1	2	3	4	5	6	7	8	9	
I	49	40	45	42	35	34	38	41	39	363
II	11	11	11	11	12	12	12	11	12	103
III	26	27	23	25	26	26	26	26	25	230
IV	11	13	6	11	10	10	8	12	11	92
V	15	17	9	23	13	15	5	16	2	115
Total	112	108	94	112	96	97	89	106	90	903

Table 3.4 Sample Size to Make Estimates of Coffe Production at the Regional Level with a CV=10%, Honduras, 1996.

Stratum	Region									Total
	1	2	3	4	5	6	7	8	9	
I	190	157	172	164	137	132	148	157	152	1409
II	43	43	45	43	48	47	47	45	47	408
III	100	106	87	96	100	99	99	100	90	877
IV	40	46	18	42	34	34	25	41	26	306
V	23	26	9	32	18	23	7	24	3	165
Total	396	378	331	377	337	335	326	367	318	3165

Tables 3.5 and 3.6 present the sample sizes required for calculating estimates for departments using expected CV's of 10 and 20% at the department level.

Table 3.5 Sample Size to Make Estimates of Coffe Production at the Departamental Level with a CV=10%, Honduras, 1996.

Dept.	Stratum					Total
	I	II	III	IV	V	
1	69	39	28	-	-	136
2	10	11	4	-	-	25
3	141	50	99	34	19	343
4	198	43	102	42	15	400
5	138	43	91	31	17	320
6	90	36	37	6	-	169
7	169	45	97	42	32	385
8	138	45	89	17	3	292
10	171	44	78	17	4	314
12	120	46	100	34	21	321
13	137	44	92	37	5	315
14	117	41	101	19	2	280
15	146	47	99	24	7	323
16	185	44	102	44	27	402
18	171	45	86	18	9	329
Total	2000	623	1205	365	161	4354

Table 3.6 Sample Size to Make Estimates of Coffe Production at the Departmental Level with a CV=20%, Honduras, 1996.

Dept.	Stratum					Total
	I	II	III	IV	V	
1	43	22	22	-	-	87
2	9	8	4	-	-	21
3	72	25	51	19	17	184
4	101	22	52	24	14	213
5	74	22	50	19	16	181
6	58	23	27	6	-	114
7	86	23	49	22	28	208
8	71	23	49	13	3	159
10	90	23	45	11	3	172
12	61	23	53	19	18	174
13	70	22	48	21	3	164
14	60	21	54	14	2	151
15	74	24	51	14	6	169
16	94	22	52	23	22	213
18	88	23	45	11	9	176
total	1051	326	652	216	141	2386

It is up to IHCAFE to evaluate its resources in terms of financial support, personnel available, transportation, and time required to do the survey in order to choose which sample design or sample size would best suit their needs. The sample sizes are minimum recommendations for each proposed scenario.

2. Recommended sample selection procedure. The following steps should be followed to select a random sample within each stratum for each region or Department. Refer to Table 3.3. In stratum I, region 1 you can see that 49 sample units were assigned to be selected from a total of 4,417 producers. To select a systematic random sample first, calculate the sampling interval.

$$\text{Sample interval} = 4,417 / 49 = 90$$

Next, using a table of random numbers, a random number between 0 and 90 is selected. If, for example, the number 22 was selected, the 22nd name on the list of the 4,417 names would be the first sample unit selected. The second name would be $22 + 90 = 112$. The third would be $112 + 90 = 202$, etc. until the 49 name is chosen. This procedure will apply to each of the 5 strata within the 9 regions.

After the samples are selected a computer file should be created for each sample.

The expansion factor will be computed for each stratum within a region using the formula:

$$FE = N/n = \text{Total Producers} / \text{Sample size}$$

3. I recommend that several samples be selected and using production data from the present database, expand and summarize the data. This total can be compared with the population or universe totals to obtain an idea of how well the samples are doing. This will give some experience with how the summary process goes and will demonstrate the distribution of the sample estimates around the population target value.

4. I recommend that a new sample be selected from the List Frame each year. The reason is that the “registry” is a dynamic list that is updated continuously as new data come into the IHCAFE office. All “registry” records that have zero coffee acreage and production should be eliminated from the List Frame sample before selecting the samples.

5. I recommend that the Area Frame sample be continued because this frame, complemented by the List Frame, gives statistical validity to the survey process. Specifically IHCAFE should:

- Continue to update the maps (cartas) as more recent editions become available;
- Improve the training of the extension agents (interviewers) that are using the maps during the surveys;
- Continue the replica rotation every 2 or 3 years;
- Transcribe the Identity and “clave” numbers so that the Area Frame sample can be matched against the complete List Frame “registry;” and
- Summarize the data for those producers that live in segments and did not match the List Frame. These tabulations will represent the “non-registry” portion of the coffee universe.

6. I recommend that a Calendar of Work be prepared that considers all the required survey activities.

- Set the date for the survey and then work backwards setting the other target dates for activities such as:
 1. Completion of the output or summary table outlines.
 2. Set the dates for training the enumerators.
 3. Set date for having all the training materials ready:

- a. Manuals
- b. Questionnaires
- c. Mapping materials
- d. Administrative materials

- Set the date for completing formulation and printing of the questionnaires.
- Determine the questionnaire content for regular and special surveys.
- Set the date for a post-survey quality control followup.

7. I recommend that the FOXPRO summary system be continued with special emphasis on check-in, coding, and editing instructions to be used before data entry begins.

- Prepare a data entry program for each specific survey.
- Prepare output or report programs for each survey.
- Prepare programs to match the identity data between the two sampling frames.

8. I recommend that the survey results be released for public use.

- Make the reports available in some form of machine media.
- Prepare historic data series, such as Table 1.1 in this report, for some of the data, that will show trends and help data users with analysis.
- Organize and maintain a library of statistical data and survey results that will facilitate searching and acquiring needed information.

9. I recommend that more emphasis be given to the data collection part of the survey process. This is where most of the survey errors are introduced! Training of the enumerators is one of the most important tasks of the Statistical Section. The enumerator must understand the questionnaire so that he can communicate its meaning to the respondent in a manner in which the respondent clearly understands what is wanted. The enumerator must be able to correctly locate the segment boundaries and interview all the producers that live inside the boundaries. If this part of the process is not correct we will be expanding and summarizing errors and incorrect data.

10. I recommend that a segment control sheet (oja de control) be prepared so that the enumerator can easily list each household in the segment, determine if it produces coffee, if so - list the identity number and "clave" number. If the household does not have a clave number then complete a questionnaire. If the household has a "clave" number then terminate the interview since this household is represented by the List Sample (registry) frame.

11. I recommend the addition of at least one more experienced person to the Statistical Section in order to more adequately accomplish the goals I have outline above, especially the training and supervision of the data collection and the subsequent quality control work.

12. PRODEPAH and IHCAFE might consider another visit by this consultant during or after the next survey in order to analyse the results of changing the List Sample frame and the implementation of the various recommendations.

SECTION IV CONCLUSION

Honduras is unique in that it has for its coffee sector the methodology established to do valid statistical surveys, and thus has the ability to make inferences about its coffee sector population with regard to such matters as:

- Area and production of coffee.
- Technologies utilized
- Social and economic characteristics of the households.
- Trends exhibited by the sector.
- Etc.

By taking advantage of and switching the source of the List Frame to the “Registry National de Productores”, while continuing with the Area Frame methodology to measure incompleteness in the List Frame, we can improve the precision of the estimates in a very economical manner.

Acknowledgements

My thanks and appreciation to Dr. Rubén Núñez and Lic. Magdalena García of the PRODEPAH office for their support and suggestions, and in IHCAFE to Ing. Franklin Osorio, Director de la División Agrícola, Ing. Carlos Fiallos and Lic. Amparo Canales de la División de Planificación for their help and cooperation in this endeavor.