PN-ABM-692
ISN 19851

ASPAP/DAI
Report No. 200

ASPAP/DSA Technical Report:

Recommended Estimation Procedures for the Rwanda National Agricultural Survey Program

Agricultural Surveys and Policy Analysis Project (ASPAP)

(Projet des Enquêtes Agricoles et Analyse des Politiques Economiques du Secteur Rural)

USAID Contract No. 696-0126

David J. Megill*
Short-Term Sampling Consultant, DAI**

Kigali, Rwanda
June 1992

*1504 Kenwood Ave., Alexandria, VA 22305, USA

**Development Alternatives, Incorporated (DAI), 7250 Woodmont Avenue, Suite 200, Bethesda, MD 20814, USA, principal institutional contractor for ASPAP, USAID Contract No. 696-0126-C-00-7777-00

TABLE OF CONTENTS

Page No.

1. Background 1

2. Finalizing Design for the National Agricultural Survey Program 1

3. Calculation of Sampling Weights 3

4. Adjustment of Weights for Missing Sample DR's 4

5. Adjustment of Weights for Non-Interview Households . . . 5

6. Optional Weight Adjustment Based on Population Projections 6

7. Survey Estimates 7

8. Calculating Variances for the Survey Estimates 8

ANNEX A. Standard Error Tables for 1990 Crop Production Survey

1. Background

The purpose of this report is to document the recommended estimation methodology for the Rwanda National Agricultural Survey data based on the new sample design, including the procedures for calculating the appropriate weights and tabulating the variances for survey estimates. These estimation procedures are based on the sample design described in David Megill's previous technical report, "Preliminary Recommendations for the Sample Design for the Rwanda National Agricultural Survey Program" (March 1992).

During Megill's previous consultancy to develop the sample design, the *Division des Statistiques Agricoles* (DSA) had not yet decided on the final number of sample *districts de recensement* (DR's) to be included in the sample. Therefore a flexible design was developed based on selecting 4 subsamples of 78 DR's each. Depending on the resources available to conduct the fieldwork, the sample for a particular agricultural survey may be based on one or more subsamples. The estimation procedures will be similar regardless of the final sample size, with the weights (expansion factors) depending on the number of subsamples included in the survey. The estimation procedures described in this report will be made flexible so they can be easily adapted to the final sample.

All work during this consultancy was again accomplished in close collaboration with Mr. Samuel Munyaneza, DSA Statistician. The invaluable assistance and support provided by Mr. Anastase Murekezi, ASPAP Director, and his staff at DSA, as well as Dr. Tardif-Douglin, Contractor Representative, Development Alternatives, Incorporated, must also be recognized.

2. Finalizing Design for the National Agricultural Survey Program

After considering the alternatives described in Megill's report on preliminary sampling recommendations, the DSA is finalizing the sampling plans for the next survey. Given the limited resources currently available for the national agricultural survey, the sample will initially consist of one subsample of 78 DR's. In following years, it will be possible to include additional subsamples in the survey as resources become available, in order to increase the precision of subnational crop production estimates.

The previous analysis comparing the crop production data from the monthly data collection procedures to the corresponding results from the weekly data collection shows potential underestimation for the production of most crops using the monthly data collection, although in many cases the differences were not statistically significant. The DSA decided to continue collecting monthly crop production data in a sample of households in order to determine whether this trend continues over time. Given the concern about the potential bias in the data collected on a monthly basis, the DSA is still studying whether to combine the crop production data based on the two procedures. The preliminary survey estimates of total crop production are currently based only on the data being collected weekly.

Each enumerator will collect data in one sample DR, given the difficulty of transportation between sample DR's. In order to maintain a reasonable enumerator workload in each sample DR, 16 sample households will be selected for the weekly crop production data collection in each DR, as well as 16 sample households for the monthly data collection procedures. This corresponds to the current enumerator workload, which appears to be working well. Following the new listing of households, a sample of four compact clusters of 12 households each will be selected in each

sample DR. Each sample compact cluster will then be split into two groups of 6 households each (4 to be interviewed and 2 as possible replacements) for the weekly and the monthly data collection. This clustering scheme ensures a higher level of correlation between the samples using the weekly and monthly data collection procedures, which should improve the analysis of the differences in the results from the two methods.

Given a sample of 16 households (with weekly data collection) in each of the 78 sample DR's, the total sample size will be 1,248 households, before accounting for noninterviews. If the DSA decides later to combine the data collected monthly with those collected weekly for tabulating the crop production estimates, the sample size would double. However, as indicated in Megill's report on sampling recommendations, the corresponding average gain in precision would only be about 8 percent, due to the large clustering effects. The decision on whether to combine the crop production data collected monthly with those collected weekly will depend on the magnitude of the bias in the monthly data, which is currently unknown. By continuing the monthly data collection procedures in a sample of households, it may be possible to estimate the approximate bias over time. If the bias is small compared to the gain in precision from the larger sample size, the data should be combined.

During this visit, Megill worked with Munyaneza on the calculation of variances for the 1990 crop production estimates, using the software package PC CARP. These results, presented in Annex A, are consistent with those produced earlier from the 1989 data. The coefficients of variation (C.V.'s) for many crop production estimates were fairly high at the *prefecture* level. Similar results can be expected from the new sample based on 78 sample DR's.

Although the sample of 78 DR's should produce reliable results at the national level, Murekezi indicated the importance of providing reasonable survey results at subnational levels at a later date when resources become available to increase the number of sample DR's. The sample of 312 sample DR's corresponding of all four subsamples should provide reliable crop production estimates at the *prefecture* level. In general, a minimum of 20 sample DR's would be required for each geographical domain in order to provide estimates of moderate reliability. This means that for a survey to produce results for each of the 32 *sous-prefectures*, a minimum sample of 640 DR's would be required. In the case of *commune*-level estimates, it would be necessary to stratify the sampling frame by *commune*, and determine the required number of sample DR's for each *commune* individually. Given that there are 145 *communes*, the average number of DR's per *commune* is about 43, although the number of DR's varies by *commune*. Since the sampling rate for DR's would approach 50 percent in many *communes*, the first stage finite population correction factor would reduce the variance considerably, thus reducing the number of sample DR's required in many *communes*. Even taking into consideration the finite population correction factors, a sample of at least 2,000 DR's would probably be required for reliable *commune*-level results.

If a project requires reliable results for a particular *sous-prefecture* or commune, it would be necessary to supplement the sample in that area with additional sample DR's. In this case, the same sampling frame used for the National Agricultural Survey Program (developed from the 1991 census data base) can be used to select the supplemental sample for that area. Once the part of the sampling frame corresponding to that area is extracted, the same sampling procedures can be used to select a larger sample of DR's with probability proportional to size (PPS).

3. Calculation of Sampling Weights

In order to expand the data from the National Agricultural Surveys to the national or domain levels, it is necessary to apply a weight (expansion factor) to the data from each survey record. The basic weight for a sample household would be equal to the inverse of its probability of selection (calculated by multiplying the probabilities at each stage of selection). Since the DR's were selected with probability proportional to size (PPS), a separate weight will have to be calculated for each DR. At the second sampling stage, a fixed number of households (16) is selected from each DR. The overall probability of selection for the sample households in each sample DR would be the following:

$$p_{hi} = \frac{n_h \times M_{hi}}{M_h} \times \frac{m_{hi}}{M'_{hi}},$$

where:

p_{hi} = sampling probability for sample households in the i-th sample DR of stratum (*prefecture*) h

n_h = number of sample DR's selected in stratum h

M_{hi} = number of households in 1991 census frame (measure of size) for the i-th sample DR in stratum h

M_h = total number of households in 1991 census frame for stratum h (cumulated measure of size)

m_{hi} = number of sample households selected in the i-th sample DR in stratum h

M'_{hi} = number of households from new listing for the i-th sample DR in stratum h

The basic weight (W_{hi}) for each sample household would therefore be the inverse of this probability, or:

$$W_{hi} = \frac{M_h}{n_h \times M_{hi}} \times \frac{M'_{hi}}{m_{hi}},$$

where W_{hi} is the weight for each sample household in the i-th sample DR of stratum h.

Note that if the number of households listed in a sample DR is equal to the corresponding number in the 1991 census frame (that is, $M'_{hi}=M_{hi}$), and if the number of households selected in each sample DR is fixed (for example, 16), the sample would be self-weighting within a stratum; that is, each sample household in a stratum would have the same weight. The weights would vary

2slightly between strata given the way the sample DR's were allocated. In practice, however, M'_h is rarely equal to M_h, so the weights will also vary by DR within stratum.

A LOTUS spreadsheet file was developed for maintaining the sampling information and calculating the weights for the sample DR's. Since the value of M'_h for each sample DR will be determined following the new listing of households, it will be necessary to update this spreadsheet after the listing operation in order to calculate the appropriate weights. It is also important to compare the value of M'_h with that of M_h for each sample DR, to make sure that the DR was properly canvassed. If there is a big difference (say, over 20 percent) between the number of households listed in the sample DR and the corresponding number in the frame, there could be a problem of interpreting the boundaries of the DR, so the supervisor should verify the listing in the field. This spreadsheet will also be updated with information on the number of completed interviews from each survey, and it will automatically adjust the weights for non-interviews, based on the adjustment factors described in the next two sections. It is designed to be flexible so that the weights can easily be revised when additional subsamples are included in the survey. A printed copy of this spreadsheet is included in Annex B, together with a list of instructions for updating the spreadsheet for each survey.

4. Adjustment of Weights for Missing Sample DR's

When an entire sample DR is not covered because of security or operational problems, it is possible to replace it with another sample DR. This procedure is based on the assumption that the replacement sample DR is representative of the original sample DR. Given the problems which make it difficult to cover the replaced sample DR, its characteristics are probably somewhat different from those of DR's which can be interviewed, thus introducing a small bias in the survey results. If a missing sample DR is not replaced, the weights would be adjusted based on the assumption that the interviewed sample DR's are representative of the missing sample DR's, resulting in a similar bias. The reason for replacing a missing sample DR is to maintain the sample size.

One way to replace a missing sample DR would be to select a sample DR from one of the subsamples not being used, since these DR's were selected using the same procedures (with probability proportional to size). In this case it would not be necessary to adjust the weights. In calculating the weight for the replacement sample DR, M_h, M'_h and m_h would correspond to the values for the new sample DR. It is recommended to select the replacement sample DR from the same *commune* as the original sample DR, when such a sample DR is available in the other subsample; if not, a sample DR from a neighboring *commune* can be used for the replacement. In this way the geographical representativeness of the sample will be maintained. In any case, the replacement DR should be selected from the same stratum (*prefecture*).

In previous surveys, the missing sample DR's were not replaced. For example, for the 1989 survey only 76 of the 78 sample DR's were covered. If the missing sample DR's are not replaced in future surveys, it will be necessary to adjust the weights. In this case, the basic sampling weight (W_h) for the households in stratum (*prefecture*) h should be multiplied by the following factor:

$$F_{1h} = \frac{n_h}{n'_h},$$

where:

F_{1h} = adjustment factor for basic weight of sample households in stratum h

n'_h = number of sample DR's actually covered in the survey in stratum h

It can be seen that the adjusted weight (W'_{hi}) can be expressed as follows:

$$W'_{hi} = W_{hi} \times F_{1h} = \frac{M_h}{n'_h \times M_{hi}} \times \frac{M'_{hi}}{m_{hi}},$$

Therefore this adjustment can be carried out by simply replacing n_h with n'_h in the weight.

5. Adjustment of Weights for Non-Interview Households

The current survey procedures involve replacing each non-interview household with another household in the same compact cluster. For the intensive (weekly) interviews, compact clusters of six households each are selected: four households in the actual sample and two for possible replacements. However, since the survey data file includes monthly crop production for the sample households throughout each crop season (6 months each), any replacement during the season results in incomplete data for the original as well as the replacement household. In order to avoid complicating the analysis of the survey data, the records for these sample households are excluded from the tabulations. One reason the non-interview households are replaced during the survey is to ensure that the replacement households are in place (and are familiar with the data collection procedures) for the next survey year; this also maintains a constant interviewer workload in each sample DR.

After the data collection for one season, the total crop production by month for the season is tabulated, excluding records for the replacement households and those being replaced (with less than 6 months of data). At the end of the agricultural year (September to August), the data from the two seasons are combined, and the annual crop production is tabulated from the records of sample households with 12 months of data. In this case, it is necessary to adjust the weights separately for the tabulations by season and for the year. In each case, the adjustment procedures would be similar. It is recommended to carry out the weight adjustment for non-interview households at the DR level.

The weight adjustment also depends on the reason for the non-interview. The weight should be adjusted to take into account sample households that were non-interviews due to refusals, temporary absence (not-at-home), etc. If a housing unit/farm is abandoned and no new household moved in, it is not a valid sampling unit. The second weight adjustment factor (for non-interview households) is defined as follows:

$$F_{2hi} = \frac{m'_{hi}}{m''_{hi}},$$

where:

F_{2hi} = weight adjustment factor for non-interview households in the i-th sample DR in stratum h

m'_{hi} = number of valid sample units (households) selected in the i-th sample DR in stratum h (excluding abandoned housing units in the sample)

m''_{hi} = number of households with data for entire reference period (included in the tabulations)

The final weight is calculated as follows:

$$W''_{hi} = W'_{hi} \times F_{2hi}$$

In order for the sample to remain representative over time, it is important to update the listing of households at least every 2 years. Since the value of M'_{hi} will change to reflect the increase (or decrease) in the number of households, the weights will be updated accordingly.

After each survey it will be necessary to update the LOTUS spreadsheet file used for maintaining the sampling information and calculating the weights, with the actual number of completed interviews. This spreadsheet will then automatically recalculate the weights based on the updated information (see Annex B).

6. Optional Weight Adjustment Based on Population Projections

Given that the weights W''_{hi} are based on the probabilities of selection, they will produce unbiased survey estimates (assuming that the listing of households is accurate). However, if it is possible to obtain good projections each year on the number of households corresponding to the frame (rural, *de facto* rural/*de jure* urban and *de jure* rural/*de facto* urban DR's) in each *prefecture*, an optional weight adjustment (in addition to F_{1h} and F_{2hi}) for the sample households in each *prefecture* could be calculated as follows:

$$F_{3h} = \frac{M'_h}{\hat{M}_h},$$

where:

M'_h = projected number of rural households (corresponding to frame) in stratum (*prefecture*) h

$$\hat{M}_h = \sum_{i=1}^{n'_h} m''_{hi} \times W''_{hi}$$

The denominator in F_{3h} corresponds to the total number of weighted households in the stratum from the survey data. This adjustment factor would be applied to the weights of all sample households in the corresponding stratum. Assuming that the projected number of rural households is accurate, this type of ratio adjustment has the effect of reducing the sampling error. The projected number of households (M'_h) would probably have to be estimated by obtaining population projections for each *prefecture* and dividing it by the average household size obtained from the 1991 Rwanda Census of Population and Housing. Since population projections at the subnational levels are especially subject to error, and the target population involves only the rural and semi-rural DR's in each *prefecture*, the DSA should be very cautious in using the F_{3h} weight adjustment factor. If the population projections by *prefecture* are not accurate, this adjustment could bias the survey results.

7. Survey Estimates

Some of the most important survey estimates will be in the form of totals, such as total crop production. The survey estimate of a total will be obtained as follows:

$$\hat{Y} = \sum_{h=1}^{10} \sum_{i=1}^{n'_h} \sum_{j=1}^{m''_{hi}} W''_{hi} \times Y_{hij},$$

where Y_{hij} = value of variable Y for the j-th sample household in the i-th sample DR in stratum h

In the case of the survey estimate of the total number of households with a particular characteristic, the variable Y may be defined as follows:

Y = 1 if the household has the particular characteristic
= 0 otherwise

Other survey estimates may be in the form of ratios, defined as follows:

$$\hat{R} = \frac{\hat{Y}}{\hat{X}},$$

where Y and X are weighted total estimates, calculated as specified above.

Survey estimates of means and proportions are special types of ratios. In the case of the survey estimate of a mean, the denominator variable X is equal to 1 for each household, so the corresponding total is equal to the sum of the weights. For the survey estimate of the proportion of households having a certain characteristic, the numerator variable is equal to 0 or 1, depending on whether a sample household has that characteristic.

8. Calculating Variances for the Survey Estimates

In the publications of the survey results it is important to include a statement on the accuracy of the survey data. In addition to presenting tables with calculated sampling errors for the most important survey estimates, the different sources of nonsampling error should be described.

The standard error, or square root of the variance, is used to measure the sampling error, although it may also include a small part of the nonsampling error. The variance estimator should take into account the different aspects of the sample design, such as the stratification and clustering. One microcomputer software package which can be used for calculating the variances for survey estimates based on a stratified two-stage sample design such as that for the National Agricultural Survey Program is PC CARP. This software can be used to calculate the variances of totals, means, proportions and other ratios, as well as regression coefficients. In the case of subpopulation estimates, the variances can be tabulated for each category of a classification variable, and it is possible to cross-classify different variables. PC CARP uses an ultimate cluster variance estimator. A copy of PC CARP is available at the DSA. This software requires an IBM-compatible microcomputer with a math coprocessor; most of the microcomputers at the DSA meet these criteria. It is recommended to use a faster machine (with a 386 chip), especially when using a large data file or when many cross-classifications are included in the analysis, given that some PC CARP runs may require considerable machine time.

The following formulas are used for calculating the ultimate cluster variance estimates for totals and ratios:

(1) Variance for the survey estimate of a total:

$$VAR(\hat{Y}) = \sum_{h=1}^{10}\left[\frac{n_h'}{(n_h'-1)} \times \sum_{i=1}^{n_h'}\left(\hat{Y}_{hi} - \frac{\hat{Y}_h}{n_h'}\right)^2\right],$$

where:

$$\hat{Y}_{hi} = \sum_{j=1}^{m''_{hi}} W''_{hi} \times Y_{hij}$$

$$\hat{Y}_{h} = \sum_{i=1}^{n'_h} \sum_{j=1}^{m''_{hi}} W''_{hi} \times Y_{hij}$$

(2) Variance for the survey estimate of a ratio:

$$VAR(\hat{R}) = VAR\left(\frac{\hat{Y}}{\hat{X}}\right) = \frac{1}{\hat{X}^2}\left[VAR(\hat{Y}) + \hat{R}^2 \times VAR(\hat{X}) - 2 \times \hat{R} \times COV(\hat{X}, \hat{Y})\right],$$

where:

$$COV(\hat{Y}, \hat{X}) = \sum_{h=1}^{10}\left[\frac{n'_h}{(n'_h - 1)} \times \sum_{i=1}^{n'_h}\left(\hat{X}_{hi} - \frac{\hat{X}_h}{n'_h}\right) \times \left(\hat{Y}_{hi} - \frac{\hat{Y}_h}{n'_h}\right)\right]$$

$\hat{Var}(Y)$ and $\hat{Var}(X)$ are variances of survey estimates of totals, calculated using the formula specified previously.

The input data files for PC CARP have to begin with the following data on the record for each sample household: (1) stratum number; (2) cluster (DR) number; and (3) weight. For the 1989 and 1990 surveys, there are 21 strata identified in the data files; these strata are described in Karen Stanecki's report on "Sample Design and Variance Estimation Procedures for the Rwanda Crop Forecasting Survey" (April 1985). It was necessary to collapse two strata (16 and 17) which had only one sample DR each. In the case of the new sample design, there are 10 strata, corresponding to the *prefectures*. These data on each record would be followed by the classification variables (for subpopulation estimates), such as *prefecture* and farm size group, and the analysis variables to be used in the tabulations, such as the production for each crop. Since PC CARP cannot transform data (by adding variables, for example), the PC CARP data input files should be generated with the data already transformed.

PC CARP was used previously to calculate the sampling errors for crop production estimates from the 1989 survey data, which were included as Annex A in Megill's previous technical report on sampling recommendations. During this visit, Megill worked again with with Munyaneza in using PC CARP for calculating the sampling errors for crop production estimates from the 1990 survey data, presented here in Annex A. Munyaneza should now have sufficient experience to generate the PC CARP input files and tabulate the sampling errors for future survey data.

The printed output of PC CARP is not easy for the data user to read and interpret, since the numbers are presented in scientific notation, the variable names cannot be more than 8 characters in length, and the categories for classification variables are only identified by their codes.

Therefore it is recommended that the PC CARP output be transformed into a publishable format by using a microcomputer spreadsheet package such as LOTUS. Such a format is illustrated by the tables in Annex A, which include the number of observations, the value of the survey estimate, sampling error, coefficient of variation (C.V.), 95 percent confidence interval and design effect (DEFF).

ANNEX A

Table 1.1. Estimates of Total Production for Selected Crops, with Corresponding Standard Errors, Coefficients of Variation, 95% Confidence Intervals and Design Effects, Based on 1990 Agricultural Production Survey - RWANDA

Number of Observations in Domain = 1,208

Crop	Estimate of Total Production (Tons)	Standard Error (Tons)	C.V. (%)	95% Confidence Interval		DEFF
				Lower Limit	Upper Limit	
Haricots Total	205,907	16,103	7.8%	174,345	237,469	4.00
Pois Total	11,036	1,200	10.9%	8,683	13,388	2.99
Sorgho	141,835	23,131	16.3%	96,498	187,172	5.82
Mais Total	95,973	18,624	19.4%	59,469	132,477	3.62
Patate Douce	819,279	61,040	7.5%	699,642	938,916	8.72
Manioc Total	264,952	27,393	10.3%	211,262	318,642	4.45
Pomme de Terre	285,033	128,688	45.1%	32,805	537,261	6.34
Banane a Cuire	602,793	58,522	9.7%	488,090	717,496	2.54
Banane a Bière	1,916,750	145,777	7.6%	1,631,027	2,202,473	4.40
Café	40,971	9,003	22.0%	23,326	58,616	9.93
Blé	7,389	2,742	37.1%	2,014	12,765	5.68
Total Number of Households	1,271,900	44,204	3.5%	1,185,260	1,358,540	-

Table 1.2. Estimates of Total Production for Selected Crops, with Corresponding Standard Errors, Coefficients of Variation, 95% Confidence Intervals and Design Effects, Based on 1990 Agricultural Production Survey - BUTARE

Number of Observations in Domain = 151

Crop	Estimate of Total Production (Tons)	Standard Error (Tons)	C.V. (%)	95% Confidence Interval		DEFF
				Lower Limit	Upper Limit	
Haricots Total	14,852	4,313	29.0%	6,398	23,306	5.34
Pois Total	890	370	41.6%	165	1,614	4.87
Sorgho	14,089	4,531	32.2%	5,209	22,969	4.36
Mais Total	2,123	1,184	55.8%	-199	4,444	7.26
Patate Douce	122,828	20,404	16.6%	82,837	162,819	3.13
Manioc Total	43,798	5,714	13.0%	32,598	54,998	0.88
Pomme de Terre	4,899	1,255	25.6%	2,439	7,359	1.88
Banane a Cuire	22,835	6,779	29.7%	9,548	36,121	4.29
Banane a Bière	194,903	42,941	22.0%	110,738	279,068	4.07
Café	3,155	1,169	37.1%	863	5,446	5.53
Blé	-	-	-	-	-	-
Total Number of Households	156,274	15,818	10.1%	125,271	187,277	1.73

Table 1.3. Estimates of Total Production for Selected Crops, with Corresponding Standard Errors, Coefficients of Variation, 95% Confidence Intervals and Design Effects, Based on 1990 Agricultural Production Survey - BYUMBA

Number of Observations in Domain = 126

Crop	Estimate of Total Production (Tons)	Standard Error (Tons)	C.V. (%)	95% Confidence Interval		DEFF
				Lower Limit	Upper Limit	
Haricots Total	37,581	8,136	21.6%	21,636	53,527	2.63
Pois Total	1,578	394	25.0%	806	2,350	1.95
Sorgho	43,172	18,867	43.7%	6,192	80,151	5.84
Mais Total	8,268	2,568	31.1%	3,235	13,300	3.00
Patate Douce	91,468	19,035	20.8%	54,158	128,777	4.02
Manioc Total	18,436	3,415	18.5%	11,742	25,129	0.67
Pomme de Terre	10,353	2,520	24.3%	5,414	15,292	1.21
Banane a Cuire	71,180	5,860	8.2%	59,695	82,664	0.40
Banane a Bière	223,214	41,196	18.5%	142,470	303,958	2.21
Café	2,283	1,108	48.5%	111	4,456	3.26
Blé	1,916	1,390	72.5%	-808	4,640	6.23
Total Number of Households	141,439	13,909	9.8%	114,178	168,700	1.46

Table 1.4. Estimates of Total Production for Selected Crops, with Corresponding Standard Errors, Coefficients of Variation, 95% Confidence Intervals and Design Effects, Based on 1990 Agricultural Production Survey - CYANGUGU

Number of Observations in Domain = 94

Crop	Estimate of Total Production (Tons)	Standard Error (Tons)	C.V. (%)	95% Confidence Interval		DEFF
				Lower Limit	Upper Limit	
Haricots Total	6,682	2,352	35.2%	2,072	11,292	6.61
Pois Total	566	245	43.2%	87	1,046	2.66
Sorgho	1,113	492	44.2%	148	2,078	1.26
Mais Total	5,153	1,431	27.8%	2,349	7,957	2.71
Patate Douce	35,591	11,620	32.7%	12,815	58,366	5.90
Manioc Total	23,382	9,440	40.4%	4,879	41,884	5.27
Pomme de Terre	3,058	1,036	33.9%	1,028	5,087	2.36
Banane a Cuire	37,056	15,308	41.3%	7,052	67,059	6.06
Banane a Bière	96,430	34,814	36.1%	28,194	164,666	5.13
Café	5,740	2,920	50.9%	17	11,463	7.86
Blé	33	33	100.0%	-32	97	0.39
Total Number of Households	90,281	15,205	16.8%	60,479	120,083	2.62

Table 1.5. Estimates of Total Production for Selected Crops, with Corresponding Standard Errors, Coefficients of Variation, 95% Confidence Intervals and Design Effects, Based on 1990 Agricultural Production Survey - GIKONGORO

Number of Observations in Domain = 91

Crop	Estimate of Total Production (Tons)	Standard Error (Tons)	C.V. (%)	95% Confidence Interval		DEFF
				Lower Limit	Upper Limit	
Haricots Total	2,905	527	18.1%	1,872	3,937	1.08
Pois Total	1,178	433	36.7%	330	2,026	2.22
Sorgho	4,616	767	16.6%	3,113	6,118	1.02
Mais Total	3,066	1,243	40.6%	629	5,503	3.44
Patate Douce	79,021	17,332	21.9%	45,051	112,991	2.96
Manioc Total	5,504	1,733	31.5%	2,107	8,901	2.57
Pomme de Terre	8,032	3,115	38.8%	1,927	14,138	2.76
Banane a Cuire	3,085	521	16.9%	2,064	4,107	0.98
Banane a Bière	50,043	4,940	9.9%	40,361	59,724	0.25
Café	1,436	400	27.9%	652	2,221	1.41
Blé	923	464	50.3%	13	1,833	2.82
Total Number of Households	100,963	7,184	7.1%	86,882	115,044	0.53

Table 1.6. Estimates of Total Production for Selected Crops, with Corresponding Standard Errors, Coefficients of Variation, 95% Confidence Intervals and Design Effects, Based on 1990 Agricultural Production Survey - GISENYI

Number of Observations in Domain = 128

Crop	Estimate of Total Production (Tons)	Standard Error (Tons)	C.V. (%)	95% Confidence Interval		DEFF
				Lower Limit	Upper Limit	
Haricots Total	9,007	3,350	37.2%	2,441	15,574	6.66
Pois Total	866	348	40.2%	184	1,548	1.91
Sorgho	1,431	879	61.5%	-293	3,154	4.33
Mais Total	30,826	16,664	54.1%	-1,834	63,486	3.63
Patate Douce	66,665	22,867	34.3%	21,846	111,484	7.10
Manioc Total	5,963	1,987	33.3%	2,069	9,858	1.93
Pomme de Terre	68,005	39,149	57.6%	-8,727	144,736	5.20
Banane a Cuire	16,598	7,775	46.8%	1,359	31,837	8.63
Banane a Bière	124,842	73,412	58.8%	-19,045	268,729	6.66
Café	8,525	7,368	86.4%	-5,916	22,965	16.38
Blé	256	167	65.1%	-71	583	1.08
Total Number of Households	124,690	15,953	12.8%	93,422	155,958	2.15

Table 1.7. Estimates of Total Production for Selected Crops, with Corresponding Standard Errors, Coefficients of Variation, 95% Confidence Intervals and Design Effects, Based on 1990 Agricultural Production Survey - GITARAMA

Number of Observations in Domain = 155

Crop	Estimate of Total Production (Tons)	Standard Error (Tons)	C.V. (%)	95% Confidence Interval		DEFF
				Lower Limit	Upper Limit	
Haricots Total	21,069	3,098	14.7%	14,998	27,141	2.41
Pois Total	998	257	25.7%	494	1,502	2.34
Sorgho	10,474	2,327	22.2%	5,912	15,035	2.52
Mais Total	3,106	794	25.6%	1,549	4,663	2.58
Patate Douce	125,313	17,453	13.9%	91,105	159,521	2.20
Manioc Total	66,675	19,666	29.5%	28,129	105,220	5.59
Pomme de Terre	4,929	1,206	24.5%	2,565	7,293	2.41
Banane a Cuire	42,993	6,660	15.5%	29,940	56,046	1.86
Banane a Bière	327,479	50,498	15.4%	228,503	426,455	2.54
Café	7,450	2,551	34.2%	2,450	12,449	3.57
Blé	687	678	98.6%	-641	2,016	1.98
Total Number of Households	164,541	10,674	6.5%	143,619	185,463	0.75

Table 1.8. Estimates of Total Production for Selected Crops, with Corresponding Standard Errors, Coefficients of Variation, 95% Confidence Intervals and Design Effects, Based on 1990 Agricultural Production Survey - KIBUNGO

Number of Observations in Domain = 95

Crop	Estimate of Total Production (Tons)	Standard Error (Tons)	C.V. (%)	95% Confidence Interval		DEFF
				Lower Limit	Upper Limit	
Haricots Total	25,936	3,682	14.2%	18,719	33,153	1.37
Pois Total	662	245	37.1%	181	1,143	1.92
Sorgho	20,079	6,327	31.5%	7,677	32,480	3.59
Mais Total	3,560	677	19.0%	2,234	4,886	1.11
Patate Douce	44,857	11,238	25.1%	22,830	66,885	3.06
Manioc Total	22,753	7,738	34.0%	7,587	37,920	3.32
Pomme de Terre	3,720	1,170	31.4%	1,427	6,012	1.46
Banane a Cuire	246,683	36,289	14.7%	175,556	317,810	1.10
Banane a Bière	333,971	35,753	10.7%	263,895	404,047	0.61
Café	4,774	1,737	36.4%	1,370	8,178	2.71
Blé	312	217	69.4%	-113	738	0.51
Total Number of Households	96,650	15,337	15.9%	66,590	126,709	2.50

Table 1.9. Estimates of Total Production for Selected Crops, with Corresponding Standard Errors, Coefficients of Variation, 95% Confidence Intervals and Design Effects, Based on 1990 Agricultural Production Survey - KIBUYE

Number of Observations in Domain = 88

Crop	Estimate of Total Production (Tons)	Standard Error (Tons)	C.V. (%)	95% Confidence Interval		DEFF
				Lower Limit	Upper Limit	
Haricots Total	6,902	545	7.9%	5,833	7,971	0.36
Pois Total	1,916	292	15.2%	1,344	2,488	0.85
Sorgho	3,455	1,890	54.7%	-249	7,158	7.51
Mais Total	18,663	5,293	28.4%	8,288	29,037	1.90
Patate Douce	57,771	6,252	10.8%	45,517	70,025	0.81
Manioc Total	7,960	2,616	32.9%	2,832	13,088	2.53
Pomme de Terre	12,197	1,236	10.1%	9,774	14,620	0.18
Banane a Cuire	5,487	3,144	57.3%	-676	11,650	7.35
Banane a Bière	39,570	15,118	38.2%	9,939	69,201	5.17
Café	175	89	50.8%	1	350	1.51
Blé	676	255	37.7%	177	1,176	2.05
Total Number of Households	91,620	3,687	4.0%	84,394	98,846	0.15

Table 1.10. Estimates of Total Production for Selected Crops, with Corresponding Standard Errors, Coefficients of Variation, 95% Confidence Intervals and Design Effects, Based on 1990 Agricultural Production Survey - KIGALI

Number of Observations in Domain = 154

Crop	Estimate of Total Production (Tons)	Standard Error (Tons)	C.V. (%)	95% Confidence Interval		DEFF
				Lower Limit	Upper Limit	
Haricots Total	55,385	9,076	16.4%	37,597	73,174	2.70
Pois Total	865	357	41.3%	165	1,565	2.48
Sorgho	27,365	7,631	27.9%	12,408	42,321	4.26
Mais Total	4,072	973	23.9%	2,164	5,979	3.28
Patate Douce	80,805	14,564	18.0%	52,259	109,351	3.85
Manioc Total	64,520	11,769	18.2%	41,454	87,587	2.79
Pomme de Terre	5,862	3,287	56.1%	-580	12,304	4.48
Banane a Cuire	126,076	38,948	30.9%	49,738	202,414	8.36
Banane a Bière	393,032	65,479	16.7%	264,693	521,371	2.72
Café	6,976	2,425	34.8%	2,223	11,730	4.68
Blé	233	117	50.4%	3	463	0.91
Total Number of Households	164,911	19,410	11.8%	126,868	202,954	2.49

Table 1.11. Estimates of Total Production for Selected Crops, with Corresponding Standard Errors, Coefficients of Variation, 95% Confidence Intervals and Design Effects, Based on 1990 Agricultural Production Survey - RUHENGERI

Number of Observations in Domain = 126

Crop	Estimate of Total Production (Tons)	Standard Error (Tons)	C.V. (%)	95% Confidence Interval		DEFF
				Lower Limit	Upper Limit	
Haricots Total	25,587	7,187	28.1%	11,500	39,674	2.92
Pois Total	1,517	664	43.8%	215	2,819	5.35
Sorgho	16,044	7,050	43.9%	2,226	29,862	4.32
Mais Total	17,138	5,249	30.6%	6,849	27,427	4.52
Patate Douce	114,961	36,149	31.4%	44,109	185,813	6.05
Manioc Total	5,962	4,243	71.2%	-2,354	14,277	7.68
Pomme de Terre	163,979	122,450	74.7%	-76,023	403,981	6.42
Banane a Cuire	30,802	12,702	41.2%	5,905	55,698	4.37
Banane a Bière	133,262	52,343	39.3%	30,671	235,853	7.72
Café	457	276	60.4%	-84	997	4.17
Blé	2,353	2,181	92.7%	-1,923	6,628	8.27
Total Number of Households	140,529	15,282	10.9%	110,577	170,481	1.77

ANNEX B

WANDA NATIONAL AGRICULTURAL SURVEY

PREADSHEET FOR MAINTAINING SAMPLING INFORMATION AND CALCULATING WEIGHTS

UBSAMPLE NO. 2

PREF.	COMM.	NAME OF COMMUNE	SECT.	NAME OF SECTEUR	DR	ZONE	R_U	TOTAL POP.	NO. OF HH'S IN FRAME FOR DR	NO. OF HH'S LISTED IN DR	NO. OF HH'S IN SAMPLE FOR DR	BASIC SAMPLING WEIGHT	NO. OF VALID SAMPLE HH'S	NO. OF COMPLETED INTERVIEWS	NON-RESP. ADJ.	FINAL WEIGHT
1	1	GISHAMVU	10	SHOLI	25	700	0	906	204							
1	5	MARABA	3	GISAKURA	21	700	0	1047	217							
1	15	NYAKIZU	7	MARABA	20	700	0	881	193							
1	18	RUNYINYA	7	MBASA	16	700	0	892	185							
1	10	MUYIRA	5	MUNYINYA	11	800	0	1341	270							
1	19	RUSATIRA	2	GAHANA	10	807	0	1200	260							
1	8	MUGUSA	3	CYAYI	33	809	0	952	229							
1	13	NTYAZO	12	NYAMURE	11	809	0	1131	274							
1	9	MUYAGA	7	NYERANZI	28	900	0	1276	310							
2	3	CYUMBA	7	MUSENDA	2	600	0	1031	216							
2	7	KIBALI	8	RUHENDA	19	600	0	1229	256							
2	12	MUKARANGE	1	BUGWE	5	600	0	989	213							
2	1	BUYOGA	3	KAVUMU	29	607	0	1009	206							
2	2	BWISIGE	3	GIHUKE	31	1106	0	1016	216							
2	16	RUTARE	7	MWENDO	23	1106	0	1180	241							
2	14	MUVUMBA	5	RUKOMO	2	1206	0	1615	353							
2	13	MURAMBI	1	GAKENKE	74	1211	0	1127	215							
2	15	NGARAMA	1	GAKOMA	17	1211	0	1530	291							
3	1	RUGARAMA	4	KIBANGIRA	3	102	0	1379	357							
3	3	GAFUNZO	4	GABIRO	3	200	0	1196	242							
3	6	GISUMA	5	GASHIRABWOBA	32	200	0	1062	196							
3	9	KARENGERA	3	GASUMO	10	201	0	1250	257							
3	10	KIRAMBO	2	GAHISI	30	300	0	1324	287							
3	7	KAGANO	10	RAMBIRA	37	302	0	1135	216							
4	4	KIVU	7	RWISHWA	13	500	0	1119	274							
4	9	MUSEBEYA	5	GIKUNGU	24	500	0	1104	260							
4	13	RWAMIKO	4	MATA	3	500	0	1348	271							
4	7	MUKO	2	CYOBE	17	507	0	1243	258							
4	1	KARAMA	1	CYANIKA	13	700	0	1319	273							
4	8	MUSANGE	5	KIBAGA	15	700	0	1177	239							
5	5	KAYOVE	8	KINUNU	44	305	0	1094	217							
5	8	NYAMYUMBA	12	RUSHUBI	3	305	0	1104	244							
5	10	RUBAVU	8	MURARA	22	403	0	1195	286							
5	1	GASEKE	3	GISEBEYA	18	500	0	1195	239							
5	4	KARAGO	7	RAMBURA	24	500	0	1225	261							
5	2	GICIYE	1	BIREMBO	40	507	0	1259	294							
5	9	RAMBA	6	NYAMPILI	6	507	0	1152	235							
5	12	SATINSYI	4	HINDIRO	10	705	0	1091	231							
6	12	NYAKABANDA	8	NGARU	37	607	0	1387	315							
6	11	NYABIKENKE	10	RUKARAGATA	34	700	0	1000	204							
6	3	KIGOMA	1	BUTARE	41	800	0	986	225							
6	6	MUKINGI	10	RUTAGARA	6	800	0	1199	239							
6	10	NTONGWE	9	NYAKABUNGO	22	800	0	1157	252							
6	13	NYAMABUYE	14	TAKWE	12	800	0	1401	279							
6	4	MASANGO	7	NYAKOGO	15	807	0	1072	213							
6	9	MUSHUBATI	6	KADUHA	23	807	0	896	176							
6	5	MUGINA	6	MUKINGA	5	809	0	1567	299							
7	11	SAKE	5	MURWA	24	1011	0	1263	298							
7	1	BIRENGA	12	SAKARA	7	1100	0	1119	210							
7	4	KIGARAMA	8	REMERA	30	1100	0	1344	262							
7	10	RUTONDE	9	SOVU	24	1100	0	851	201							
7	3	KAYONZA	7	RWINKWAVU	8	1112	0	1459	331							
7	9	RUSUMO	2	GISENYI	60	1112	0	1091	230							
7	9	RUSUMO	9	NYAMUGARI	73	1112	0	1440	384							

SUBSAMPLE NO. 2 (Continued)

PREF.	COMM.	NAME OF COMMUNE	SECT.	NAME OF SECTEUR	DR	ZONE	R_U	TOTAL POP.	NO. OF HH'S IN FRAME FOR DR	NO. OF HH'S LISTED IN DR	NO. OF HH'S IN SAMPLE FOR DR	BASIC SAMPLING WEIGHT	NO. OF VALID SAMPLE HH'S	NO. OF COMPLETED INTERVIEWS	NON-RESP. ADJ.	FINAL WEIGHT
8	1	BWAKIRA	3	MUGUNDA	9	705	0	1548	317							
8	2	GISHYITA	8	NGOMA	7	305	0	1775	372							
8	3	GISOVU	5	KAVUMU	23	503	0	1255	277							
8	4	GITESI	12	RURAGWE	29	503	0	922	197							
8	8	RUTSIRO	3	GASOVU	35	503	0	1182	236							
8	5	KIVUMU	8	NGOBAGOBA	31	507	0	1224	265							
9	1	BICUMBI	9	MWULIRE	71	1110	0	1070	208							
9	9	MUGAMBAZI	8	NTYABA	29	700	0	1279	257							
9	14	RUTONGO	5	KABUYE	27	700	0	1824	413							
9	16	TARE	1	BUMBA	7	706	0	1046	205							
9	7	KANZENZE	11	MWOGO	2	910	0	1295	290							
9	11	NGENDA	8	NYARUGENGE	67	1009	0	1368	309							
9	3	GASHORA	8	RILIMA	34	1011	0	1420	285							
9	5	GIKORO	2	DUHA	35	1100	0	1290	259							
9	12	RUBUNGO	4	KINYINYA	2	1100	0	1393	324							
11	7	KINIGI	3	GIHORA	19	400	0	1058	214							
11	6	KIGOMBE	3	GAHONDOGO	107	400	1	600	129							
11	5	KIDAHO	5	CYANIKA	18	406	0	1227	263							
11	12	NYAKINAMA	4	KABERE III	21	504	0	929	209							
11	1	BUTARO	4	KAYANGE	41	600	0	1259	302							
11	3	CYERU	12	RUHOMBO	28	600	0	1059	206							
11	16	RUHONDO	4	MUKONO	19	607	0	1057	222							
11	4	GATONDE	3	GAHANGA	19	706	0	1152	234							
11	15	NYARUTOVU	3	GASHENYI	47	706	0	1122	244							

SAMPLING PARAMETERS BY STRATUM

STRATUM (PREF.)	NO. OF HH'S IN 1991 CENSUS FRAME	NO. OF DR'S SELECTED IN STRATUM
1	160550	9
2	162939	9
3	103818	6
4	100119	4
5	150163	8
6	170789	9
7	136202	7
8	97042	6
9&10	193398	9
11	161191	9

Instructions for Updating Spreadsheet for Calculating Weights

1. The first eight columns in the spreadsheet involve information from the sampling frame and should not change. Therefore these columns in the spreadsheet should be protected.

2. The column for "No. of Hh's Listed in DR" should be updated after the new listing. In the report this value is referred to as M'_{hi}. This information should also be updated at least once every two years.

3. The column for "No. of Hh's in Sample for DR" should be 16 based on the current sampling plan, but this may vary by survey. In the report it is referred to as m_{hi}. If the crop production data collected weekly is combined with the monthly data, then m_{hi} would be equal to 32 based on the current design.

4. The column for "Basic Sampling Weight" appears in the spreadsheet as a formula. If the spreadsheet is changed, it is important to verify that the formula for each cell corresponds to the equation for W_{hi} specified in the report.

5. The column for "No. of Valid Sample Hhs." should be updated after each survey. In the report it is referred to as m'_{hi}. Although its value is generally equal to m_{hi}, it excludes any sample housing unit/farm which is abandoned.

6. The column for "No. of Completed Interviews" should be updated after each survey. In the report it is referred to as m''_{hi}. This is the number of households in the sample DR with completed interviews (with production data for all months) which are actually used in the tabulations.

7. The column for "Non-Response Adjustment" appears in the spreadsheet as a formula. It corresponds to F_{2hi} in the report.

8. The column for "Final Weight" also appears in the spreadsheet as a formula. It corresponds to W''_{hi} in the report. This is the weight which should be attached to all the records for the sample households in the DR.

9. In the table on "Sampling Parameters by Stratum" below the sample DR's in the spreadsheet, the column for "No. of Hhs. in 1991 Census Frame" should not change. The column for "No. of DR's Selected in Stratum" should correspond to n'_h in the report, that is, the number of sample DR's actually covered by the survey in the stratum. In this way the weight will automatically be adjusted for missing sample DR's.

10. In order to assign the weights to the survey records, the spreadsheet can be used to obtain an ASCII file that contains geographical identification of the DR's and the corresponding final weights. An SPSS program can then be developed to assign the weights to the records by DR number.

11. The information for the DR's in the 3 remaining subsamples are included in the spreadsheet, below the information corresponding to Subsample 2 (selected for the first survey). If any additional subsample is used for a survey, the spreadsheet can be revised to include them in the weight calculations. In this

case, all the formulas can be copied for the additional sample DR's; for the basic sampling weights (W_{hi}), the formulas should be copied from the corresponding cells for that stratum in Subsample 2. Again, it is important to verify these formulas after revising the spreadsheet. The column for "No. of DR's Selected In Stratum" should be updated to reflect the total number of DR's in all subsamples included in the survey.

12. The spreadsheet with the final weights for each survey should be saved under a new name which identifies the survey and the year. These files should also be backed up, of course.