1973 Crop Production in Upper Volta

by

Fred B. Warren

Mathematical Statistician

Statistical Reporting Service

U.S. Department of Agriculture

cooperating with

U.S. Agency for International Development

May 1974

Index

Page

I. Summary..... 1

II. 1973 Crop Production Survey

A. Survey Design..... 3

B. Execution..... 3

C. Summarization..... 4

D. Evaluation..... 4

E. Estimates..... 5

III. Sampling on Future Surveys

A. Sample sizes for crop surveys in Upper Volta..... 8

B. Sampling for specific crops..... 11

C. Selecting villages with equal probabilities, or with probabilities proportional to the size of the village..... 11

D. Random or systematic sampling..... 12

IV. Recommendations..... 12

1973 Crop Production in Upper Volta

I. SUMMARY

This report includes a discussion of a small crop production survey in Upper Volta in 1973 as conceived, as executed, and as summarized with recommendations for future surveys.

The intention of this survey was that it would supply U.S.AID with estimates of moderate precision of the production of millet and sorghum in Upper Volta in 1973. This information would have been useful in the administration of the drought relief program to that country.

My recommendations[1/] for this survey were that Upper Volta would need notice of U.S.AID approval for the project by October 3, 1973, and the necessary equipment should be in Ouagadougou by October 10, so that there would be a minimum amount of time in which to organize a survey which would have to be underway by October 15. Actual notification of approval was not received in Ouagadougou until October 17, one day before I arrived with the equipment required for the survey.

Since notification of approval had not been received earlier, Upper Volta had made no preparations for the survey. The two weeks (October 18 to November 3), I had in Upper Volta were used in assisting the Upper Volta statistical office to prepare for the survey. By the time the survey was started, all crops had been harvested. The agents conducting the survey did attempt to obtain data on the total amount of material harvested from the sample fields. However, this data was for grain in the head, in non-standard containers, and was judged not useable.

The survey plan was to measure and sample the fields belonging to two randomly selected households in each of two systematically selected villages in each O.R.D. However, no data was received from two O.R.D.'s and from only one village in each of two other O.R.D.'s. For the area that was included in the survey, the relative sampling errors computed for the areas planted to millet (11.5 percent) and to all crops (21.9 percent) were within the expected range of reliability.

I was also asked to review a draft copy of a request for U.S.AID assistance in setting up a program of agricultural statistics in Upper Volta. My comments on this were sent to D. L. Atwell, Ouagadougou, on April 24, and are largely repeated in this report. Briefly, I would state:

1/ "A Millet and Sorghum Crop Estimating Program for Niger and Upper Volta," p. 21, Fred B. Warren, SRS-USDA, cooperating with U.S.AID.

1. Upper Volta needs everything they asked for,

2. They also need technical assistance, training in the area of sampling and of summarizing survey results, and

3. I am concerned as to if the absolutely essential vehicles could be obtained in time for a survey in 1974.

II. 1973 CROP PRODUCTION SURVEY

A. Survey Design

The survey plan called for a total sample size of 22 villages with two households to be selected from each village. Two villages were selected with equal probabilities from each O.R.D. The villages were selected systematically rather than as a simple random sample to insure that the sample villages would not be located near to each other.

The usual two-stage procedure for selecting sample households is to (1) list all of the households in the village, and then (2) select two households at random and with equal probabilities from that number.

Because of the amount of time required to list all of the households and because of the very limited amount of time available for the survey, this procedure was modified to allow villages of more than 40 households to be divided roughly into quartiers of from 20 to 30 households. The expectation was that this could be accomplished as a result of an interview with the village chief to determine the limits of the village area and a visual inspection of the village area to see where the households were located. One quartier would then be selected at random and the households in that quartier enumerated. Information obtained during this enumeration included the names, age, and sex, and matital status of all members of each household, the pricipal activity of the head of the household, and whether any members of the household had any fields of millet, sorgho, maize, rice, or fonio. Two agricultural households were then selected at random from that quartier. All fields belonging to each of the two households were then measured and were to have been sampled for yield.

B. Execution

The millet harvest started around October 1. Unfortunately, approval of the survey by AID was not received in Upper Volta until October 17, one day before I arrived with the equipment needed for the field measurements. Therefore, no preparations for the survey had been made at that time. It required an additional 8 days to get the equipment through customs. The head of the statistical office, Mr. Garey, was then free to discuss preparations for the survey.

All crops in all parts of the country had been harvested before the Ministry of Agriculture was able to find a vehicle so that Mr. Garey could get out to the O.R.D.'s to instruct the agents who were to collect the data on this survey. Therefore, the only useable data from this survey was for field areas. I did not receive any data for the O.R.D.'s. of Diebougou-Gaoua and Fada-N'Gourma. Also, data from only one village rather than two, was received from the O.R.D.'s of Koupela and of the Sahel.

Many of the agents did attempt to salvage some yield information by asking about the production from the fields. The answers obtained were always in terms of volume of some type of container-panniers, sacs, greneries, or charretees. Many also attempted to obtain a weight for volume conversion factor for the particular container, but neglected to state if the weight was for shelled or unshelled grain. If unshelled, then a second conversion to shelled grain would also be needed.

C. Summarization

The summarization of survey data was limited to the computation of field areas, estimation of areas planted and harvested to the cereal crops enumerated on this survey, and variance components for areas planted to major crops. The summarization was carried out on a remote access terminal in the U.S. Department of Agriculture in Washington, D.C. The computer program used were adapted from those written earlier for the Niger Crop Production Survey.

D. Evaluation

1. The forms prepared by Mr. Garey for recording the survey forms were quite good. I was particularly impressed by the field measurement forms which require the enumerator to take compass directions from each end of each side of the field. (The visée and contrevisée). It also provides space for measuring diagonals across the field. Whether for this or other reasons, less than 5 percent of the fields measurement had substantial errors of measurement and they were easily corrected.

   I would suggest that a numerical identification code be part of of each document. Such a code should identify the document as to O.R.D., village, household, and field. These codes would be essential if they are to go to computer processing.

2. My intention was that villages would be subdivided into quartiers only if there were at least 35 or 40 households in the village. This intention obviously was twisted in the transmission of instructions, and in the understanding of the agents who acted as enumerators. In only 6 of the 16 villages did the enumerators actually select quartiers having as many as 20 households.

3. Subject to the crippling limitations of timing and transportation to which this survey was subjected, I feel that the data collection generally was well done.

E. Estimates

This survey was intended to produce estimates of area and production of crops for the entire country of Upper Volta. Because no data was obtained from the O.R.D.'s of Fada N'Gourma and of Diebougou-Gaoua, this was not possible. Therefore, the results presented here will pertain only to the other nine O.R.D.'s. Two different estimators were used. The first is the unbiased direct-expansion estimator which uses the reciprocal of the probabilities of selection as the expansion factors in constructing the estimate. For example, if

M is the total number of villages in an O.R.D., and

m is the number of sample villages selected,

then if the villages are selected with equal probabilities (as we did on this survey), the probability of selecting any one village is:

$$P_v = m/M$$

Similarly, if N_i is the total number of agricultural households in the i'th sample village, and

n_i is the number of households selected for the sample with equal probabilities, then

$$P_i = n_i/N_i$$

Then if X_{ij} is the total area of millet (or of any other estimable item) for the j'th sample household in the i'th sample village, the direct expansion estimate (X) of that item for the O.R.D. is computed as:

$$X = (1/P_v)\left(\sum_{i=1}^{m} (1/P_i) \sum_{j=1}^{n} X_{ij}\right)$$

$$= (M/m)\left(\sum_{i=1}^{m} (N_i/n_i) \sum_{j=1}^{n} X_{ij}\right)$$

Although unbiased, the direct expansion estimate can produce misleading results if the sample villages are much different in size from the average of all villages in the O.R.D. (Table 1) A way to adjust for the size of the villages is to divide the direct-expansion estimate (X) by the ratio (R) of the average size of the sample villages to the average size of all villages. Algebraically, the ratio estimate (X_r) is computed as:

$$X_r = X/R.$$

Table 1--Number and average size of villages, by O.R.D.'s, Upper Volta

	Villages	Population[1/]	Population per Village	Population per sample Village	Ratio
	(no.)	(no.)	(no.)	(no.)	(R)
Ouagadougou	1088	960,475[2/]	885	602	.681
Koudougou	670	899,361	1,342	1,409	1.048
Diebougou-Gaoua	1095	396,474	362	168	.464
Sud-Ouest	197	176,189	895	1,102	1.232
Bobo-Droulasso	448	440,122	984	274	.279
Volta-Noire	834	575,906	692	211	.305
Yatenga	667	604,990	906	1,615	1.781
Kaya	674	700,404	1,041	270	.264
Fada N'Gourma	729	348,330	478	522	1.092
Koupela	484	368,002	760	586	.771
Sahel	459	3/	3/	3/	3/

1/ Population figures from last census

2/ Excludes Ouagadougou commune

3/ Data was not available

Estimates of the total area planted to major crops in Upper Volta in 1973 and the corresponding relative standard errors are given in Table 2. The only appreciable differences between the two sets of estimates is for the area planted to millet.

Table 2--Direct expansion and ratio estimates of areas planted to major crops in Upper Volta in 1973, with relative standard errors.

Item	Direct Expansion		Ratio Expansion	
	Estimate	Relative Standard Error	Estimate	Relative Standard Error
	(000 Hectares)	(Pct.)	(000 Hectares)	(Pct.)
All Crops	1862.4	21.9	1541.3	24.2
Millet				
- pure stands	1041.5	9.5	778.9	12.5
- interplanted	79.1	89.9	42.3	84.6
- total	1120.6	11.5	821.2	12.5
Sorgho				
- pure stands	667.4	50.9	617.5	46.4
- interplanted	122.0	65.4	141.9	48.9
- total	789.3	36.7	759.4	35.0
Maize - total	94.1	49.0	95.9	47.7

III. SAMPLING ON FUTURE SURVEYS

A. Sample sizes for crop surveys in Upper Volta

In two-stage cluster sampling such as used for the crop surveys in Upper Volta and in Niger, the optimum number ($\bar{n}$) of sample households per sample village can be estimated as:

$$(1) \quad \bar{n} = \sqrt{\frac{C_v}{C_h} \cdot \frac{S_h^2}{S_v^2}} \quad , \text{ where}$$

C_v is the cost of travelling to and listing the households in a sample village.

C_h is the cost associated with each sample household. This would include measuring the fields, locating the sample plots, and harvesting and weighing the production from the sample plots.

S_h^2 is the between household (within village) component of variance, and,

S_v^2 is the between village (within O.R.D.) component of variance.

The variance components listed in Table 3 were estimated from the 1973 survey data for Upper Volta. Since I have no knowledge of either the actual or relative values of C_h and C_v, I have assumed that they are equal. In this event, the ratio (C_v/C_h) is one so that equation (1) reduces to:

$$(2) \quad \bar{n} = \sqrt{S_h^2 / S_v^2} \, .$$

Equation (2) was used in preparing line 4 of Table 3. As the statistical office in Upper Volta determines more realistic values for C_h and C_v, they can compute revised values for $\bar{n}$ by multiplying the indicated values by the factor $\sqrt{C_v/C_h}$.

The indicated optimum number of sample households per village normally will not be a nice even number such as 1, 2, or 3. Also, if optimum numbers are computed for several different items, they probably will not agree. In this example, the optimum number of households per village could be as low as 1.06 (if estimating sorgho) or as high as 2.97 (for estimating millet). Since it would not be practicable to select a different number of households for millet than for sorgho, a likely compromise would be to select 2 sample households per village.

The total number of sample villages to be selected in each O.R.D. can be limited either by the number of people and other resources that are available, or by the desire to produce estimates which have at least a minimum degree of accuracy as measured by the ratio of the estimate (X) to its sampling error (S). (This ratio is often called the relative sampling error. If expressed as a percentage it would be called the coefficient of variation or C.V.). Given a number of sample households per village (n), and a sample of village which had been selected with probabilities proportionate to the number of agricultural households in each, the minimum number of sample villages required to produce an estimate (X) having a given sampling error (S) can be estimated as:

$$m = (X/S)\sqrt{(S_v^2 + S_h^2/n)}$$

Table 3--Variance components and optimum sample allocations, Upper Volta, 1973.

Item	Total Cropland	Area of Sorgho	Area of Millet	Area of Maize
Average Area (X) per Household (hectares)	3.27	1.99	1.51	0.54
Between household Component of variance (S_h^2)	3.74	2.67	2.03	2.13
Between village Component of variance (S_v^2)	1.57	2.39	0.23	-.46
Optimum number of sample households per village ($\sqrt{S_h^2 / S_v^2}$)	1.54	1.06	2.97	1/
Number of villages required for a 10 percent relative sample error	37.3	124.1	40.1	1/
Probable average relative sampling error with a sample of 37 villages an 2 households per village (pct.)	9.3	15.9	12.1	23.7

1/ Not computed due to a negative component of variance.

B. Sampling for Specific Crops

The present survey procedure of selecting 2 households at random from the entire village works well for estimating quantities of crops which are grown by everyone in the village. It does not work as well for other crops simply because the growers of those crops may not be selected. For sufficiently large samples, the growers of the minor crops would be selected in approximately the proper proportions so that the survey results will still be unbiased, but the sampling errors will be large.

One way to reduce the sampling errors for the minor crops would be to draw separate samples of households using the information recorded on Fiche 1. For example, if 14 of the 20 households in a sample village had any millet, 13 had any sorgho, 10 had any maize, 6 had any fonio and 4 had any rice, then the normal procedure would be to randomly select 2 households from the 20 at random, regardless of which crops were grown by them. Then the within-village expansion factor would be $\frac{N}{n} = 20/2 = 10$ for each crop grown by the two selected households. In this example, it is unlikely that any of the 4 households growing rice would be selected.

An alternative procedure would be to:

1. Select two households from the total number in the village but do not include any rice fields that these two households may have. The normal procedures will be used for all other crops.

2. Select 2 households that grow rice from the total number of households growing rice. These two households will be used only for these rice fields. The expansion factor for rice in this village then would be $4/2 = 2.0$.

3. The above procedure could be extended to other crops or groups of crops as desired.

C. Selecting villages with equal probabilities, or with probabilities proportional to the size of the village.

The advantages and rationals for selecting sample villages with probabilities proportionate to size rather than with equal probabilities was documented in my companion report on Niger[1/]. It would appear that the potential gains from sampling with probabilities proportional to size rather than with equal probabilities would be at least at large in Upper Volta as in Niger. The basic unit of size to be used should be total population excluding any urban areas such as Ouagadougou commune and any other known centers of non-agricultural activity.

1/ "Crop Production in Niger (1973)," by Fred B. Warren, SRS-USDA, cooperatin with U.S.AID.

D. Random or systematic sampling

The particular advantages and disadvantages of systematic sampling as compared with simple random sampling are these.

Advantages

1. When properly done, systematic sampling will insure that the sample units (e.g. villages), are distributed more evenly over the area being sampled.

2. In certain conditions, particularly when sampling with equal probabilities, a systematic sampling can be selected more easily than a simple random sample.

3. If a village is apt to be more like nearby villages than like those which are farther away, then systematic sampling will often result in estimates which have smaller sampling errors than would be obtained from a simple random sample of the same size. Items to be considered for crop surveys would include the types of crop grown and potential yields per hectare.

Disadvantages

1. If the purpose of the systematic sample is to obtain a more even distribution of villages over an administrative area or of households over a village area, than the villages (or households) must be arranged in some order.

2. The sampling errors computed from a systematic sample will be over-estimates of the actual sampling error.

In Upper Volta, lists of villages arranged by small administrative areas are already available. Also, the lists of households in the sample village would normally start at a central point and proceed from concession to concession in order of proximity so that the list of households (Fiche 1) is already in a satisfactory order for a systematic sample.

Therefore, it seems that selecting a systematic sample of villages within O.R.D.'s would be perferred for crop surveys in Upper Volta. The desirability of the systematic sampling procedure for selecting sample households within village would depend upon an evaluation by the statistical office in Upper Volta as to whether or not the yields and types of crops grown by adjacent households would be more alike than for households which were widely separated in the village.

IV. RECOMMENDATIONS

The Ministry of Agriculture in Upper Volta has established its intention of developing an on-going program of agricultural statistics by training 225 local agents in the techniques of data collection. However, the lack of money for the necessary equipment and supplies has kept them from actually doing anything.

I have reviewed a draft copy of a request for U.S.AID money. This money would be used to purchase necessary equipment and to defray operating costs over a two-year period. At the end of that time, the expectation embodied in the proposal was that the budget of the Ministry of Agriculture would have been increased by enough so that the program could continue.

Classes of items included in the request were vehicles and gasoline for them, survey equipment for measuring fields, various types of office equipment including calculators, and salaries. I would agree that all of these items would be required for the successful operation of a program of collecting agricultural statistics.

The type of statistical program to be implemented was not spelled out in any detail. Generally, it would collect data on area and production of crops, market prices and marketing practices, and the use of various types of improved forming methods. I would assume that the foundation of such a program would be a national sample survey to estimate area and production of various crops. This survey would be similar to the 1973 survey, but with the following recommended changes.

1. The sample size would be increased to the point where the sampling error of the total area under cultivation would be less than 5 percent. This would probably require at least 200 villages. However, they should not attempt to survey more than half that number the first year.

2. The survey should start in July, just after the crops were planted. The field measurements should be summarized immediately so that national and regional estimates of the areas seeded to different crops could be published no later than September.

3. Sample villages should be selected systematically within regions, with probabilities proportional to the best estimate of the size of the individual villages.

4. Sample households should be selected from the entire village, not from randomly selected quartiers.

5. That "special purpose" samples be selected for minor crops, in those sample villages which had any of them.

6. That once a sample household has been selected, that it be used for 2 or 3 years in succession. This would permit the use of a ratio to last year estimate which would give a much better indi-

cation of year to year changes than the direct expansion estimate described in Section II E.

This annual survey could also be used in obtain information on numbers of livestock (donkey sheep, goats, chickens, cattle, etc.) in the agricultural sector.

I have an uneasy feeling that the proposed program of surveys will generate more data than can be summarized by hand with the number of people that I think will be available. The possible use of the IBM 360-25 computer in Ouagadougou should be considered. If it can be used, the programs I wrote for the 1973 data could probably be adapted without too much trouble.

I would also recommend that any U.S.AID grants of money for materials, etc., be accompanied by technical assistance in terms of a series of 2 or 3 weeks TDY's. A possible schedule for such a series would be:

1. June 1973 - 1 sampling statistician
   1 programmer statistician

   2 weeks in Upper Volta to assist with sample selection and to assist with plans for tabulation of the survey results.

2. September 1974 - 1 sampling statistician
   1 programmer statistician

   3 weeks in Upper Volta to:

   a). evaluate and assist in the summarization of the field area data,

   b). assist with final plans for summarizing the yield and production data, and

   c). observe the collection of yield information in different parts of the country.

3. March 1975 - 1 sampling statistician
   1 programmer statistician

   2 weeks in Upper Volta to reivew experiences of the 1974 and to develop changes for future surveys.

4. April 1975 - 2 survey statisticians

   4 weeks in Upper Volta to assist in the conduct of an in-country training program for permanent staff. This program would be designed to the fundamentals of elementary statistics and to develop the basic concepts of sampling applications to agricultural data systems.

5. If this program of further technical assistance is requested from SRS, it is suggested that an additional 2 man-months be included for preparation and other work which would need to be done in Washington.