Complex Survey Design & Other Challenges

Zambian Agricultural Surveys

Dr. Cynthia Donovan
Dept. of Agricultural Economics, MSU

Presentation to
Stats 825, Sample Surveys
November 28, 2007

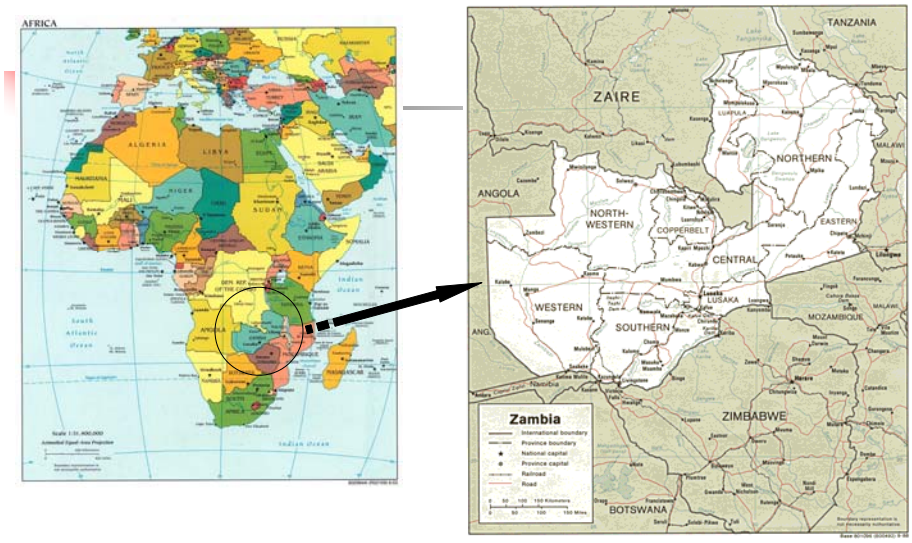
USAID FROM THE AMERICAN PEOPLE

MICHIGAN STATE UNIVERSITY

Content

- Objectives of survey in Zambia
- Factors that go into survey sampling generally
- Choices made in the Zambian case
- Effects of those choices on usefulness of data
- Problems found and possible solutions

Zambia: Slightly bigger than Texas



Zambia

- 9 provinces
- 72 districts
- Census 1990 (most recent)
- Out of 8 million people (1.7 million households), approx. 55% are in agricultural sector (about 930,000 households)

Post Harvest Survey Objectives

- Nationwide representative agricultural survey for small-scale farmers
  - total production of key crops
  - total area under key crops
  - prices received by farmers
  - inputs used by farmers
  - services received by farmers
  - other aspects

What is needed

- Determining sampling strategy
- Methods for measurement of key aspects
- Implementation strategy, given constraints
- Final analytical challenges

Factors in Sample Size

- Characteristics to evaluate & their variability
- Extrapolation to a population
- Precision/confidence in the estimates
- Effective sample size for key aspects
- Probability of non-response due to inaccessibility, refusal, etc.
- Controlling non-sampling errors
- Costs: available funds
- Time constraints

Precision and sample size

Example: From population of 5000 farm hhs, what proportion cultivate rice if a sample size of n=50 households found 10 rice farmers?

A 95% Confidence Interval for prop. would be:

$P +/- 2*[(1-f)*(pq)/n]^{1/2} = (0.087, 0.312)$

Thus, 8.7% to 31.2% of the households cultivate rice, very wide estimate

To narrow CI band: larger sample or lower CL

Strategies in sample selection

- Simple Random Sampling
- Clustering
- Stratifying
- Complex Sampling (combinations)

Strategies in sample selection

- Simple Random Sampling
  - Ease of analysis
  - Large sample to get precision in some cases
  - Can be high cost
  - Can take more time

Strategies in sample selection

- Clustering
  - Lowers cost
    - Travel
    - Supervision
  - Decreases time
  - Loss of precision in estimates
  - Good when elements w/in cluster are heterogeneous

Strategies in sample selection

- Stratification
  - Can decrease variability
  - Differential sampling for distinct sub-pops
  - May decrease costs

Doing both: Complex sampling with complex results on precision

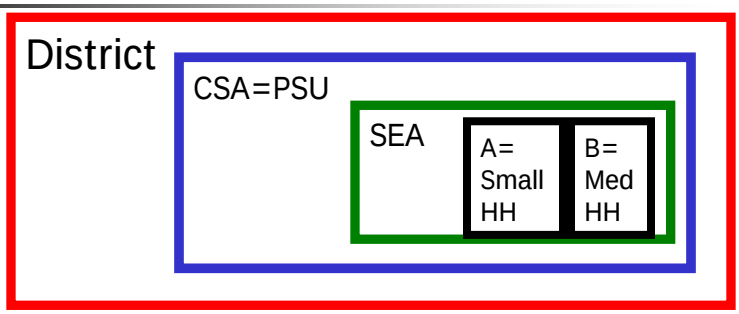
Stratification

- Need to have adequate numbers in subpopulations of interest:
  - Medium scale farm households
  - Less than 10 % of population
  - Need to “over-sample”

Design Effects

- DEFF: Ratio of variance of the estimate obtained from the complex sampling to the variance obtained from a simple random sample of same size
- High DEFF means important to incorporate sampling design into estimations
- Relationship btw. weights and DEFF
- Software example: SPSS vs STATA

Post Harvest Survey 1997/98: Stratified multi-stage survey



CSA= Census Supervisory Area
PSU= Primary Sampling Unit
SEA= Standard Enumeration Area

Note: PHS 997/98 is used here as the example.
Sampling and results are different for more recent years.

Sampling:

Stage	Sampling Unit	Stratification
1	CSA	District- Province
2	SEA	
3	Household	Category A/ Category B

CSA=Census Supervisory Area
SEA=Standard Enumeration Area:

Problems found

- Sample size/strategy good for some crops, but not for others (localized, specialized crops)

Problems found (cont.)

Crop	Production	C.V. (%)	C.I.: Upper	C.I.: Lower	Design Effect
Maize	623,580	4.64	566,923	680,237	2.32
Sorghum	23,598	14.19	17,033	30,163	5.50
Rice	7,536	22.54	4,207	10,865	4.36

Very wide CI and high DEFF for some crops and activities

Problems found

- Sample size/strategy good for some crops, but not for others (localized, specialized crops)
- Sample size not distributed for provincial level analysis

Table 1. Number of Sample SEAs and Households for the 1997/98 Post-Harvest Survey, by Province

Geographic Area	Number of Sample SEAs Selected	Number of Sample SEAs Covered	Number of Sample Households Selected	Number of Sample Households Contacted
Zambia	379	378	7550	6336
Central	40	40	800	674
Copperbelt	24	24	480	349
Eastern	68	68	1360	1197
Luapula	48	47	939	775
Lusaka	14	14	280	252
Northern	72	72	1440	1190
North Western	27	27	531	423
Southern	49	49	980	828
Western	37	37	740	648

Problems found

- Sample size/strategy good for some crops, but not for others (localized, specialized crops)
- Sample size not distributed for provincial level analysis

- Non-response with possible bias due to access
- Inclusion of nonagricultural households (rural sample)

- 750,000 sq km
- Bad roads
- Need low cost
- Some areas low pop density

Non-response and non-ag

Overall "non-response" rate:
16% of households in listing

(up to 27% in Copperbelt; 20% in Northwestern
17% each in Luapula and Northern; 10% in Lusaka)

Variety of reasons:
hh not found (3%)
hh moved away (10%)
hh dissolved (2%)
hh refused to participate (1%)

Non-ag hhs in sample: 182 hhs (3%)

Problems found

- Sample size/strategy good for some crops, but not for others (localized, specialized crops)
- Sample size not distributed for provincial level analysis
- Non-response with possible bias due to access
- Inclusion of nonagricultural households (rural sample)

- Stratification on landholding size could be more effective

Table 2. Distribution of Sample Households by Farm Size and Listing Category

Farm Size Group	Listing Category		
	Total	Cat. A	Cat. B
Total	6,154	5,559	595
0 ha.	120	117	3
0.01-0.99 ha.	2,639	2,614	25
1.00-1.99 has.	1,894	1,819	75
2.00-2.99 has.	695	623	72
3.00-3.99 has.	297	224	73
4.00-4.99 has.	160	89	71
5.00-9.99 has.	303	66	237
10+ has.	46	7	39

Problems found

- Sample size/strategy good for some crops, but not for others (localized, specialized crops)
- Sample size not distributed for provincial level analysis
- Non-response with possible bias due to access
- Inclusion of nonagricultural households (rural sample)
- Stratification on landholding size could be more effective

Solutions

- More money, can increase overall sample size
- Exclude non-ag rural hhs from sample frame
- Stratification for localized crops/activities
- More efficient landholding stratification
- Separate surveys/samples for different characteristics
- Change sampling for more disaggregated analysis

Conclusion

- Use statistics to design the sampling
- Know the basic characteristics that you want to evaluate before you design the sampling
- Use prior surveys and small select surveys to help determine sample size
- Combine statistical and financial criteria to determine feasible and yet acceptable options
  - Different sampling costs
  - Different sampling efficiency
  - Cures for access issues

Thanks to

Ministry of Agriculture and Cooperatives (MACO)

Zambian Central Statistical Office (CSO)

US Bureau of the Census Intl. Programs

And

USAID
FROM THE AMERICAN PEOPLE