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:

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

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
  - Ease of analysis
  - Large sample to get precision in some cases
  - Can be high cost
  - Can take more time
- Clustering
- Stratifying
- Complex Sampling (combinations)
Strategies in sample selection

- Clustering
  - Lowers cost
    - Travel
    - Supervision
  - Decreases time
  - Loss of precision in estimates
  - Good when elements within 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

District

CSA=PSU

SEA

A= Small HH
B= Med HH

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

Sampling:

<table>
<thead>
<tr>
<th>Stage</th>
<th>Sampling Unit</th>
<th>Stratification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>CSA</td>
<td>District- Province</td>
</tr>
<tr>
<td>2</td>
<td>SEA</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Household</td>
<td>Category A/ Category B</td>
</tr>
</tbody>
</table>

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.)

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

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

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

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

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