Chapter VI

SAMPLING, AND HOW IT MIGHT BE IMPROVED

It became clear in Chapter III that there were some sample biases in all villages, but that the number of biased variables and the intensity of the biases varied considerably from village to village. On Tai proved to have the least biased sample, while that for Buak Khang was the most biased. Of course, Buak Khang was not a true sample. However, the case of Buak Khang is instructive as more than just a bad sample. It also shows that although two villages may be located that close together, and appear that similar, they may still prove to be quite different, so that generalization from one village to another is very dangerous.

After on Tai as the best sample came Ban Chung, Thap Nam, and Khayai in that order. It is interesting that the villages which depended mainly on mono-cropping (On Tai and Ban Chung) should have the samples with the least amount of bias. In Thap Nam, which was mainly an agricultural village, but which was intensively involved in the raising of livestock, the sample showed more bias than in the two villages dependent on mono-cropping, but not as much as in Khayai. As for Khayai, which was involved in both agricultural and non-agricultural activities more or less equally, the sample proved poor, even when considered separately for the agricultural and non-agricultural sectors.

As has already been mentioned several times earlier, in the cases of Ban Chung and Thap Nam, the samples of households with land holdings proved more representative of their populations than did the whole village samples. Also in Khayai, dividing the sample and population into land-holding and landless households made the comparison between population and sample more relevant and much more fruitful. From this it would appear that greater care should be taken as to what population is to be sampled, and what variables are to be used.

The assumption that rural households are agricultural households is often not valid. Nor is the assumption that agricultural households in Thailand are mainly involved in crop production always valid. The first step in selecting a sample for the intensive study of labour utilization is to establish what population is to be studied. Depending on the intended use of the results, definite decisions should be made as to what constitutes the population for the study, whether it is the population of an entire village with a diversity of activities, the population of the agricultural households in that village, the population of the rice-growing households in that village, etc.

Although the intended use of the results are the most important determinant of the population to be sampled, the constraint imposed by the size of the sample should also be considered.

There is a maximum size sample possible for any village depending
on the manpower available to collect the data. If the maximum
sample size were limited to 15 households, in a large village of
several hundred households with wide diversification of economic
pursuits, this would mean that a sample of the whole village which
attempted to represent the diversity would be weak, if not of
questionable validity. In such a case, a decision might have to
be made to restrict the population, resulting in more homogeneity,
and a greater likelihood of being able to represent it in a small
sample.

Once a population has been decided upon, the problem becomes one
of deciding what variables should be used. In the following
section, several variables relevant to labour utilization will
be discussed briefly.

Probably the variable which appears most obvious in the study of
rural manpower utilization is the land holding situation. A house-
hold which has a land holding (whether the land be owned, rented
in, or both, or of some other tenure) will use its labour force
differently from a household which does not hold any land. However,
the minimum size of holding required to employ and support a farm
family must not be overlooked here.[1] Households with holdings

---

1. In Ayutthaya, Fuhs (1974) estimates that a holding of appro-
   ximately 25 rai is necessary to provide employment and income
   for the average farm family, while Amyot (1974) considers the
   minimal holding to be 15 rai. For villages in changwat
   Chiang Mai, the minimal holding is somewhat smaller, and in
   many areas of the Northeast, somewhat larger.

smaller than the minimal holding may utilize their available labour
in a way that is more similar to landless households. For sampling
purposes, then, many of the villages to be found in rural Thailand
should be divided into three groups of households, those without
land, those with less than a minimal holding, and those with a
holding that is minimal or larger. If the appropriate minimal
holding in known, sampling with these three groups in mind is rela-
tively simple. If not, then probably the whole range of land
holdings should be considered.

Presumably, sampling on a basis of land holdings (including holding
size) gives a reasonably good basis for determining in each village
situation the amount of time spent on crops, and the amount of time
available for other activities. This is based on the usually
correct assumption that crop production is the central concern of
farmers. It is probably because the sampling for the five villages
discussed in this study was done primarily on the basis of land
holdings that there was so little bias on variables related to crop
production (at least, when only the land-holding households were
considered.)

Crop-production, however, is only one of the activities pursued
by rural households, and especially where mono-cropping is practised,
other activities become important, as there is a long slack period
between crops, as well as a shorter slack period while the crop is

growing. In the land-holding sample households in the three villages in Ayutthaya (Ban Chung, Khayai, and Thap Nam), Fuhs (1974) reported that work on farms amounted to 46% of the total time available,[2] and the time spent on crops amounted to 32%, 13%, and 16% respectively of the total time available. (Crops and agricultural support together accounted for 37%, 16%, and 21% respectively of total time available, or for 40.6%, 16.1%, and 21.3% respectively of the total time worked.)

These data make it clear that other activities beside crop production are pursued even by households with land holdings, and that they are of some importance in those households. The more important activities for both households with and without land are livestock raising, cottage industry, off-farm employment (both agricultural and non-agricultural), off-farm work pursued own-account, and domestic work. In the five villages which have been discussed in the present study, it was in connection with these types of activities that sample biases occurred.

The present problem, then, is how to go about sampling in such a way as to take these activities into account. Many variables are involved, including the number in the household labour force, the number of children younger than eleven years, the number of

---

2. For the definition of "total time available" see Fuhs (1974)

children eleven years or older studying full-time, the number of old
or handicapped persons, the industries of main and subsidiary occu-
pation and employment statuses of labour force members in these
occupations, the participation of the labour force in own-account
non-farm work, and in both agricultural and non-agricultural
employment, the numbers of livestock raised and sold, etc. Some
of these variables are applicable to all villages, while others
are only important in certain villages. Obviously taking all of
these variables into consideration at all times is a very unyieldy
procedure and quite impractical.

The present investigator would suggest that one set of these
variables is less important than the others, and can be left uncon-
sidered. This is the set related to household structure. Three of
these variables (namely the number of persons in the household,
the number too old to work, and the number physically or mentally
handicapped), failed to show biases in the five villages studied.
Three other variables in this group, the number of persons in the
labour force, the number of children younger than eleven years,
and the number of children eleven years or older engaged in full-time
study or training, did show biases in some of the villages studied.
Of these three, the most important variable for labour utilization
is the size of the labour force. The number of children younger
than eleven, and the number eleven or older studying usually reflect
the household labour force situation, the latter because usually

children ... are only allowed to study if they are not needed in the labour force, and the former because most of the labour force is in the child-bearing and child-raising age-groups, and although there may be a period in most households when most or all of the children are grown and at home, and do not yet have children of their own, this is a temporary state of affairs. In addition, these two variables can be considered relatively unimportant with regard to labour utilization, as their greatest impact is on domestic work.

Although the size of the labour force is an important variable, as it is the basis of the availability of labour, special measures to control it may not be necessary either. Except in areas where there are very few employment opportunities, the size of the labour force is related to opportunities for gainful employment both inside and outside the household. If the household does not provide adequate opportunities, labour force members look for employment outside the household. If this means going any distance, they usually migrate, seasonally or permanently. Thus, if samples are adequately selected on the basis of the types of activities pursued by the labour force, bias with respect to labour force size may cease to be a problem.

The sampling problem now seems to be reduced to how to consider the different types of activities proportionately to their occurrence and relative importance in the village. Using the variables from the Socio-Economic Profile Schedule as a basis for doing this seems unduly complicated.

The idea of using just the main and subsidiary occupations of the household head deserves consideration. The problem with this approach is that the household head's occupation may not necessarily reflect the occupations of the rest of the household labour force. In fact, the distribution of occupation of household heads is usually much more diversified than that of the labour force as a whole (see Amyot, 1974, and other data available at CUSSRI).

A second approach would be to consider the occupations of all the labour force members. If this were to be done, some kind of an index or technique for summarizing each household would have to be worked out, for the basic sampling unit is the household, not the individual labour force member. It seems complicated to collect such detailed data and then have to summarize it in order to sample.

The present investigator has come to the conclusion that a new approach not directly dependent on the Socio-Economic Profile Schedule should be taken for the selection of the intensively studied household sample. One suitable approach will be suggested. The basis for this approach is the assumption that the important activities pursued by each household in the population can be identified by asking an adult member what the main occupation and the most important subsidiary occupation of the household as a whole are (irrespective of the occupations of any particular labour force member). As well as these two occupations, the relative size of

each enterprise should be found out. (How size is determined will
depend, of course, on the nature of each enterprise.)

The first step after the collection of the data would be to stratify
the population on the basis of main occupation. If the important
main occupations were activities such as crop production, or live-
stock raising, and there were a diversity of sizes, each stratum
would then be sub-stratified on the basis of size of enterprise.
Then, within these sub-strata, the subsidiary occupation would be
considered, and if one were lucky enough to have that homogeneous
a population, the size of the enterprise once again. It is the
present investigator's guess that usually the major divisions would
occur for main occupation and the size of enterprise for main occu-
pation, and that the subsidiary occupation would only serve to avoid
the selection of atypical households within the already determined
strata for main occupation, and size of enterprise of main occupation.

For the collection of the data on the main and subsidiary occupations
of each household, the intention here is not to use the classifica-
tion system of the Socio-Economic Profile Schedule (see Table V-2
in Appendix A for this classification), but something more similar
to the activities of the labour utilization data. A classification
such as the following would be most appropriate:

1. Agricultural work on-farm - Crops

2. Agricultural work on-farm - Livestock

3. Agricultural employment

4. Cottage Industry - with one category for each type of
   cottage industry in the village.

5. Other (than cottage industry) own-account non-farm work

6. Non-agricultural employment - with one category for
   each important type of work done.

This could be expanded further where required or appropriate, by
dividing crops into rice and rubber, for example, or livestock
into the different kinds.

Several modifications of the stratification system are also
possible. If the main interest is in labour utilized for agri-
culture in an area where there are both land-holding and landless
households, the population might be reduced to include only house-
holds with land holdings. (The problem posed earlier about mini-
mal holdings would have to be taken into consideration here.)
Or, in a case like Khayai, with a split between important acti-
vities, the population might be divided into two on this basis,
and each section sampled from separately. (In this case, sampling
would not necessarily be strictly proportional to the frequency
of occurrence of the two activities in the population. Within
each activity, however, sampling would proceed as already outlined.)
The major reason for doing this would be to get a large enough
sample of each section to be able to make meaningful comparisons
between them, when sampling on a basis of the total population
would mean too small a sample for one section for the data to be

considered representative. What modifications might be required depends on the central concern of the study, and the characteristics of the villages being included in it.

A sampling scheme such as the one outlined above has two advantages for a labour utilization study. First, it samples from among the types of activities actually being pursued, and is therefore directly related to the utilization of labour. Secondly, it is relatively easy to collect the necessary data, and to go through the mechanics of sampling. However, it does require adding some time to the schedule for such a study, as it requires data collection from the whole village before sampling can be done, and therefore before data collection on the households to be studied intensively can be started.

Following is a hypothetical example of a village to be sampled from. Suppose this village contains 85 households, and the plan is to take a 15 household sample (18%). (Although determining the number of cases in the sample, rather than the percentage, is somewhat unusual in sociological surveys, it is preferable in this instance, because the nature of the study requires not a one-shot interview, but daily data collection over one year. 15 households is, under most conditions, an optimimum number for adequate coverage by one researcher.) With a 15 household sample in a village of 85 households, each sample household represents 5.7 village households, or 6.66% of the village household population.

Suppose further that the data for industry of main occupation show there are 4 industries pursued in the village: 1. Agriculture - crop production; 2. Agriculture - livestock raising; 3. Cottage Industry (brick-making); and 4. Off-farm employment. Of these four, suppose that livestock raising is unimportant, accounting for only 2% of the village households (i. e. as industry of <u>main</u> occupation).

The first step is to stratify on industry of main occupation. Suppose the population distribution is:

Crop production	55%
Livestock raising	2%
Brick-making	31%
Employment, off-farm	12%

The number of sample households are then assigned proportionately. This can be done either by using the percentages for the population distribution and calculating the number of sample households by rounding (i.e. N = 55/100 X 15 = 8.25, or <u>8</u>), or by using the percent of population households represented by each sample household (6.66%) and calculating how many multiples of 6.66 are required to approximate the population frequency (i.e. N = 55/6.66 - 8.25, or <u>8</u>).

The result of either of these approaches is:

Industry of Main Occupation of Households -categories-	Population Distribution (percent)	Number of sample households	Sample Distribution (percent)
Crop production	55	8	54
Livestock raising	2	0	0
Brick-making	31	5	33
Employment, off-farm	12	2	13
Total	100	15	100
	(N=85)		(N=15)

The percentages for the population and sample distributions in each category are not equal, but are close to equal. Greater equality is not possible, as the basic sampling unit is the household, and it is not possible to use fractions of households. The differences between the two distributions should never exceed one-half of the percentage of population households represented by 1 sample household (i.e. 3.33% in this case, as $\frac{1}{2}$ of 6.66% = 3.33%). As livestock raising accounts for only 2% of the population, and that is less than 3.33%, it is not included in the sample.

The second step is to stratify each of these categories on size of enterprise. The crop production category will be used as an example. Suppose that in the given range of sizes of holding

in the population, 20 rai can be considered an appropriate
dividing point between small and large holdings. Also suppose
that the population distribution of the size of operational
holding within the sub-group with crop production as industry of
main occupation is:

less than 20 rai   60%

20 rai or more   40%

The resulting sub-stratification would then be:

Size of Operational Holding - categories -	Population Distribution (percent)	Number of Sample Households	Sample Distribution (percent)
less than 20 rai	60	5	63
20 rai or more	40	3	37
Total	100	8	100
	(N=47)		(N= 8)

A second example of Step II stratification is the category for
off-farm employment. Suppose that there is a limited number of
types of employment available for off-farm employment, and that
the conditions and pay for all of these are very similar. Stra-
tification on size of enterprise is not relevant, therefore it
need   not be done.

Step III is to stratify each of the sub-strata from Step II on
industry of subsidiary occupation, in the same way as Step I

was done. Step IV is to stratify the sub-strata from Step III on size of enterprise of subsidiary occupation for those sub- -strata for which a further breakdown is relevant, if any.

The results of the complete stratification of the hypothetical village are shown in **Figure I.** Sampling is then done on the basis of random selection of the required household(s) from each of the final sub-strata. In practice, probably the most convenient way to do this would be to put the data on small cards, with one card per household. The various breakdowns could then be easily done by sorting the cards into piles, the result of the whole procedure being a pile of cards for each of the final sub-strata, each pile containing all households belonging to that sub-stratum.

The only way to be sure how effective this new sampling approach might be is to use it, and then study the results. However, the fact that the new approach takes into considera- tion the major economic pursuits of the village households, rather than just the land available for crop production, should mean that it would result in samples less biased on variables related to labour utilization (as well as on incomes and expen- ditures).

As mentioned several times earlier, another source of bias will still be present. Once households have been selected for the sample, it will still be necessary to ascertain whether the

selected households are willing to co-operate with the research assistants for the one year period. If it appears that any household may not be willing to co-operate, it would be better to replace that household with another one (randomly selected from the sub-stratum from which the original household was selected) than to try to include the unco-operative household only to have to drop it from the sample after it is too late to replace it.

More complex stratifications than the one suggested above are, of course, possible. The present investigator thinks, however, that it is preferable to use a scheme such as the one suggested, rather than to make the procedure unduly complex. The scheme suggested requires brief preliminary interviews of all households. Other schemes would require even more preliminary data collection, and much more complicated procedures for stratifying on the variables to be used.

...terprise of — Percent of Sub-population in this Category	Number of Sample House-holds	Step III: additional assumptions	Category Variable: Industry of Subsidiary Occupation	Percent of Sub-Population in this Category	Number of Sample House-holds	Step IV: additional assumptions	Category Variable: Size of Enterprise of Subsidiary Occupation	Percent of Sub-Population in this Category	Number of Sample House-holds	Result: Number of Sample House-holds	Percent of Total Population
60	5		Livestock raising	18	1						
			Cottage Industry	50	2	wide range of sizes necessitates further subdivision	less than 1,500baht	43	1	1	5.9
							1,500 baht or more	57	1	1	7.1
							Subtotal (StepIII)	100	2	1	9.4
			Employment, off-farm	32	2						
			Subtotal (step II)	100	5					2	10.5
40	3		Livestock	21	1	-	-	-	-	-	-
			Cottage Industry	42	1	-	-	-	-	1	4.6
			Employment, off-farm	39	1	-	-	-	-	1	9.2
			Subtotal (step II)	100	3	-	-	-	-	1	8.5
100	8					-	-	-	-	-	-
-	-	-	-	-	-	-	-	-	-	0	2.0
65	3		Crop production	56	2					2	11.2
			Employment, off-farm	44	1					1	8.8
			Subtotal (step II)	100	3	-	-	-	-	-	-
35	2		No subsidiary occu-pation	60	1					1	6.5
			Crop production	40	1					1	4.3
			Subtotal (step II)	100	2	-	-	-	-	-	-
100	5										
100	2		No Subsidiary occupation	90	2					2	10.8
			Cottage industry	10	0					0	1.2
			Subtotal (step II)	100	2						

Figure I. Stratification System for Sampling from a Hypothetical Village.

Basic Assumptions	Step I: Stratify on Industry of Main Occupation			Step II: Stratify on S... Main Occupat...	
	Category Variable: Industry of Main Occupation	Percent of Population in this Category	Number of Sample households	Additional assumptions	Category Varia... Size of Enterp... of Main Occupa...
The village has 85 households. Project resources make possible a 15 household sample (i.e. 18% of the village population). That means each sample household theo-retically should represent 6 village households, or 6.7% of the population of village households	Crop production	55	8	assume that 20 rai is an appropriate dividing point	Operational la... holding less t... 20 rai
					Operational l... holding 20 ra... or more
					Subtotal(step
	Livestock raising	2	0		-
The frequency of occurrance of the main and subsi-diary occupations and the sizes of enterprises of these are as found in the table	Cottage Industry	31	5	assume a wide range of sizes of enterprise, with 1,500baht annual income an appropriate dividing point between large & small enter-prises	Annual income cottage indust... less than 1,50 baht
					Annual income Cottage indust... 1,500 baht or Subtotal(step
	Employment, off-farm	12	2	assume a very narrow range of size of enterprise, so that no sub-division is necessary	
	Total	100	15	-	-