

## Sample Design for Nam Ha and Ninh Binh Survey

The 1995 Vietnam Longitudinal Survey was designed as a probability sample survey of about **2000** households (weighted) of **10** communes in Nam Ha and Ninh Binh provinces. According to data from the 1989 census, the two provinces consisted of 471 communes with 773,308 households. With these data, we have designed a two-stage, stratified random sample selection procedure. To ensure that every household will have a known probability of selection, the sampling procedures are designed so that each household in both provinces will have the same sample fraction  $f$ , where  $f = 1,600 / 773,308 =$  **0.00207** (unweighted sample size divided by total number of households of Nam Ha and Ninh Binh provinces). However, because we wish to have a minimum sample of four urban communes, we have adjusted the number of households sampled in the four urban communes. The sample fraction for urban Nam Dinh is adjusted 2 times greater than the general sample fraction and the sample fraction for two urban communes in Phu Ly and Ninh Binh is 10 times greater than the general sample fraction. Accordingly, the number of interviewed households is estimated to be **1,855**.

### **Sampling procedures:**

#### Commune stratification

Because socioeconomic conditions are different in urban and rural areas, and distances from highways differ as well, all communes in the two survey provinces are stratified into four groups.

Group 1: Urban communes of three provincial city/towns - the city of Nam Dinh, the town of Ha Nam, and the town of Ninh Binh. This group consists of 23 urban communes with 46,534 households. For each city/town, select 1 urban commune.

Group 2: Communes that are within about 3 km from national highway 1 or inter-provincial high ways 10 or 21. The group has 104 communes with 156,853 households. Select 2 communes.

Group 3: Communes that are within 3 - 10 km from national highway 1 or inter-provincial highways 10 or 21. This group has 107 communes with 179,222 households.  
Select 2 communes.

Group 4: This is the largest group, consisting of 237 communes and located more than 10 km away from national highways 1 or inter provincial highways 10 or 21. There are 390,699 households in this group. Select 3 communes.

### Commune selection

Based on estimates for transportation and other expenses of survey management and on the geographic distribution of the two provinces, a total of 10 communes will be sampled. Three urban communes make up for group 1; group 2 and 3 will have two communes; group 4 has 3 communes. Sampled communes are chosen separately for each group with a probability proportion to number of households (this can be considered as equivalent with a probability proportion to size of population). Give  $f_1$  as probability of commune selections, then:

$$f_1 = n * H_c / \sum H_c \quad (1)$$

Where  $n$  = number of communes to be sampled,  
 $H_c$  = number of households of commune  $c$ ,  
 $\sum H_c$  = total number of households of a strata.

For a given group of communes, the selection procedures are as follows:

- Find the selection **interval - I**:  $I = \sum H_c / n$
- Find a random number  $R_1$  between 0 and I. Because we need to choose  $n$  communes, we must have  $n$  random numbers; the systematic series of random numbers after  $R_1$  is:

$$R_i = R_1 + i * I \quad i = 2, \dots, n$$

- From a sampling frame (list of communes for each group) with a number of households from each commune, create a column of cumulative sum, find two communes such that  $R_i$  falls in between. If  $R_i$  falls between communes  $k$  and  $k+1$ , commune  $k+1$  will be a sampled commune.

The following 10 communes were selected for the survey: Truong Thi (Nam Dinh City), Minh Khai (Ha Nam Town), Van Giang (Ninh Binh Town), Loc Hoa (Nam Dinh City),

Trinh Xa (Binh Luc-Nam Ha), Thanh Nghi (Thanh Liem-Nam Ha), Nam Duong (Nam Ninh-Nam Ha), Xuan Thanh (Xuan Thuy-Nam Ha), Hai Dong (Hai Hau-Nam Ha), Khanh Mau (Yen Khanh-Ninh Binh).

### Household selection

For each household to have the same opportunity to be selected,  $f_2$  (probability of household selection for a sampled commune) will be:  $f_2 = f / f_1$ . The number of households to be interviewed ( $n_c$ ) for a commune is computed by multiplying the number of households of commune ( $H_c$ ) by  $f_2$ .

Because:

$$\begin{aligned} n_c &= H_c * f_2 \\ &= H_c * f / f_1 \\ &= (H_c * f) : (n * H_c / \sum H_c) \\ &= f * \sum H_c / n \end{aligned} \quad (2),$$

$n_c$  is depended on sampling fraction ( $f$ ) and number of sampled communes ( $n$ ) for each strata. In each strata,  $f$  and  $n$  are constant, so sample size (number of surveyed households) for every commune will be the same.

During the fieldwork, the households to be interviewed ( $n_c$ ) will be selected by the following procedures:

- Create randomly a list of households of the commune. Count the number of commune households at the time of survey in 1995:  $H_c$ .
- Define the sampling interval  $I = H_c / n_c$ . Take a random number from 1 to  $I$  for the first sampled household in the list, then for every  $I$  households, select another household until we have  $n_c$  households.

Probabilities  $f_1$ ,  $f_2$  (weighted for urban communes) and sample size for each sampled commune are as follows:

*Sample Design for Nam Ha and Ninh Binh Survey*

<b>ID number</b>	<b>Name of commune</b>	$f_1$	$f_2$	Nc	Adjusted sample size
250113	P. Truong Thi, Nam Dinh	.061510	.06731	2276	153
250202	P. Minh Khai, Ha Nam	.399549	.05181	1593	83
280102	P. Van Giang, Ninh Binh	.268530	.07709	1489	115
250122	X. Loc Hoa, Nam Dinh	.025081	.08253	1967	162
250712	X. Thanh Nghi, Thanh Liem	.026254	.07885	2039	162
250606	X. Trinh xa, Binh Luc	.013659	.15155	1224	185
251017	X. Nam Duong, Nam Ninh	.024249	.08536	2173	185
251106	X. Xuân Thanh, Xuân Thuy	.009667	.21412	1295	270
251311	X. Hai Dong, Hai Hau	.013215	.15664	1721	270
280706	X. Khanh Mau, Yen Khanh	.010681	.19380	1391	270

Note: Total sampled households 1855 calculated from the list of households in the 1989 census. Using the data from 1995 can make the sample fraction about 15% smaller in reality.