

## **II. SAMPLE DESIGN AND ESTIMATION**

### **1. Sample frame**

The sample of the 2010 labour force survey is the two-stage stratified sample, represents for provincial level. Each centrally governed province, city constitutes a main stratum with two sub-stratums of urban areas and rural areas. The sample frame is sample 15% enumeration areas of the 2009 Population and Housing Census.

### **2. Determination of sample size and sample allocation**

To ensure an obtainment of sample estimates representative for provinces, the sample was allocated inversely to population size, each province had a sample size of about 60 enumeration areas with an average of 32 households per enumeration area. Hanoi and Ho Chi Minh City are the two most populous and complicated cities, so they had a selection of 74 enumeration areas (Annex 2). For sample size, average 2,000 households per province per period. Each period of the 2010 labour force survey was conducted in 3,890 sample enumeration areas in entire country. The General Statistics Office selected directly enumeration areas and informed Provincial Statistics Offices.

The sample of the survey is the stratified sample, each province, city constitutes a main stratum (63 main strata) with two sub-stratums of “urban” and “rural” areas. Sample allocation for each stratum was conducted by the system selection method. Sample enumeration areas were selected in the preparatory step, the Department of Population and Labour Statistics conducted to select the first level system as enumeration area, the second level: randomly selected 32 households from the household listing of enumeration areas.

### **3. Sample estimation**

#### ***3.1 Sample estimation for one enumeration period***

*The general weight can be calculated from the following probabilities/weights:*

- 1) Design weights (basic weights): depend on probability;

- 2) Coefficients used to adjust weights due to variations in the number of households or variations in the total number of enumeration areas because of lost but not replaced by other ones;
- 3) Coefficients used to adjust weights according to structure of the population under study (self-weighting).

*Notations:*

- $W_{1hji}$  - Design weights (basic weights) of  $j^{th}$  EA,  $h^{th}$  stratum;
- $W_{2hji}$  - Coefficients used to adjust weights due to variations in the number of households (persons);
- $W_{3hji}$  - Coefficients used to adjust weights according to the average number of households in EA of  $h^{th}$  stratum;
- $W_{4hji}$  - Coefficients used to adjust weights due to variations in the number of enumerator areas;
- $W_{5hji}$  - Coefficients used to adjust weights according to structure of the universal;
- $W_{hji}$  - Weights for the number of households or male/female population of  $j^{th}$  EA,  $h^{th}$  stratum.

*Determination of basic weights*

According to the sample design, the labour force survey was conducted on a sample size almost equal throughout all 63 provinces/cities. Each province/city conducted interviews to about 24,000 persons.

Suppose  $a_h$  is the number of EAs selected in the  $h^{th}$  stratum and  $N_h$  is the total number of EAs of  $h^{th}$  stratum. Because the sample was selected independently for each stratum according to the systematic random sampling method, so the basic selection probability is calculated by the formula:  $P_{1hji} = \frac{a_h}{N_h}$  and the basic weight (design weight) of the  $j^{th}$  EA,  $h^{th}$  stratum is the inverse of the selection probability, which was calculated as follows:

$$W_{1hji} = \frac{1}{P_{1hji}} = \frac{N_h}{a_h} \approx \frac{M_h}{\sum m_{hj}}$$

In which,  $M_h$  is the total number of households (persons) of  $h^{th}$  stratum and  $\sum m_{hj}$  is the total number of households (persons) of all selected EAs in the  $h^{th}$  stratum.

*Determination of coefficients used to adjust weights due to variations in the number of households (persons) and the number of EAs*

- a) Adjustment of weights due to variation in the number of households (persons):

Suppose  $m_{hj}$  is the total number of households (persons) at the time that  $j^{th}$  EA,  $h^{th}$  stratum carried out its listing and  $m_{hj}^*$  is the total number of households (persons) at the time that  $j^{th}$  EA,  $h^{th}$  stratum conducted its enumeration. Because EAs in each stratum is selected with equal probability and were selected from the 2009 Census sample frame, and coefficients used to adjust the weights due to variation in the number of households (persons) will be calculates by formula:

$$W_{2hji} = \frac{1}{P_{2hji}} = \frac{m_{hj}}{m_{hj}^*}$$

As the 2009 Census's enumerator areas were created with unequal size, so it is necessary to determine coefficients used to adjust the size of each of 2009 Census EAs to be equal to the average size of EAs in that stratum. Suppose  $\bar{m}_{hj}$  is the average number of households (persons) of EAs in  $h^{th}$  stratum, then the coefficients used to adjust the weights due to variation in the number of households (persons) will be calculates by formula:

$$W_{3hji} = \frac{1}{P_{3hji}} = \frac{\bar{m}_{hj}}{m_{hj}}$$

- b) Adjustment of weights due to variation in the number of EAs:

It was stipulated in the 1/4/2010 survey that if a selected EA was found to be destroyed or cleared away during the update of maps and lists, the nit would be replaced by a neighbor to keep the total number of selected EAs in the stratum unchanged, hence:

$$W_{4hji} = \frac{1}{P_{4hji}} = 1$$

***Determination of coefficients used to adjust weights by structure of the universal (weights by proportion of the studied population)***

The annual average population is estimated by urban/rural and sex for each of 63 provinces/cities, so it could be weighed according to proportions of urban and rural population, and male and female population. Suppose  $m_{hji}^*$  is the total number of households (population by male/female) recorded by enumerators for  $j^{th}$  EA,  $h^{th}$  stratum ;  $m_{hji}^{*'} is the total number of households (population by male/female) adjusted according to proportions of urban/rural population and proportions of male/female population of  $j^{th}$  EA,  $h^{th}$  stratum, then it is calculated by formula:$

$$m_{hji}^{*'} = m_{hj}^* \times \frac{M_{hi}^*}{M_h^*}$$

Where:

- $m_{hj}^*$  Total number of households (population by male/female) provided by the sample survey for  $j^{th}$  EA,  $h^{th}$  stratum;
- $M_{hi}^*$  Total number of households (population by male/female) by urban/rural area estimated as of 1/4/2010 (or 1/10/2010) for  $h^{th}$  stratum ( $i = 1 - \text{urban}; i = 2 - \text{rural}$ )
- $M_h^*$  Total number of households (persons) estimated as of 1/4/2010 (or 1/10/2010) for  $h^{th}$  stratum.

Coefficient used to adjust weights according to structure of the population (households) estimated as of 1/4/2010 (or 1/10/2010) are determined as follows:

$$W_{5hji} = \frac{1}{P_{5hji}} = \frac{m_{hji}^*}{m_{hji}^{*'}} \times \frac{M_h^*}{M_h} = \frac{m_{hj}^*}{m_{hji}^{*'}} \times \frac{M_{hi}^*}{M_h^*} \times \frac{M_h^*}{M_h} = \frac{m_{hj}^*}{m_{hji}^{*'}} \times \frac{M_{hi}^*}{M_h}$$

As the sample allocation is not proportional to the population under study, so sample weights will be calculated to be used in analyzing the results of the survey to ensure an actual representation of the sample. The weight for each household (or population ( $i$ )) in urban area ( $i$ ) of  $j^{th}$  EA,  $h^{th}$  stratum is the inverse of selection probability:

$$W_{hji} = 1/P_{hji} = W_{1hji} \times W_{2hji} \times W_{3hji} \times W_{4hji} \times W_{5hji}$$

$$W_{hji} = 1/P_{hji} = \frac{M_h}{\sum m_{hj}} \times \frac{m_{hj}}{m_{hj}^*} \times \frac{\bar{m}_{hj}}{m_{hj}} \times \frac{m_{hj}^*}{m_{hji}^*} \times \frac{M_{hi}^*}{M_h} = \frac{\bar{m}_{hj}}{\sum m_{hj}} \times \frac{M_{hi}^*}{m_{hji}^*}$$

The Department of Population and Labour Statistics cooperated closely with the Statistics Informatics Center of Zone I to programme, calculate specifically and check accurately above weights for all of 3890 enumeration areas.

### ***3.2 Average sample estimation for enumeration periods***

The 2010 Labour force survey was conducted at 2 time-points (period 1: 1/4/2010 and period 2: 1/10/2010). To weight the data for 2010, it is needed to the general weight for both of the two enumeration periods as the following formular:

$$W_{hji}^* = \frac{W_{hji}^1 + W_{hji}^2}{2}$$

Where:

- $W_{hji}^*$  - The year weight for households or population (male/female) of the enumeration area  $j$  of the stratum  $h$ .
- $W_{hji}^1$  - The period 1's weight for households or population (male/female) of the enumeration area  $j$  of the stratum  $h$ .
- $W_{hji}^2$  - The period 2's weight for households or population (male/female) of the enumeration area  $j$  of the stratum  $h$ .