

# **Sampling Design of Socio-Economic Survey of Cambodia (SESC) 1996**

***Gloria A. Cubinar***  
***Sampling Expert***

## ***1. Sample Design Summary***

The sampling design for the 1996 Socio-Economic Survey of Cambodia (SESC) is a two-stage stratified sample, and it is a probability sample for a coverage area of 87.26 percent of Cambodian villages. Unlike in the previous survey where the population were stratified into three (3) strata namely, Phnom Penh, Other Urban and Other Rural, this year's SESC's number of strata will be increased to come up with the estimates for the following provinces: Banteay Meanchey, Battambang, Kampong Thom, Pursat, Ratanak Kiri, Siem Reap and Svay Rieng. Thus, the sample will now consist of ten (10) strata. These are: Banteay Meanchey, Battambang, Kampong Thom, Phnom Penh, Pursat, Ratanak Kiri, Siem Reap, Svay Rieng, Other Urban and Other Rural.

For each survey round, three hundred seventy five (375) primary sampling units (PSU's) or a total of seven hundred fifty (750) PSU's for the two rounds, defined as villages will be selected using the systematic sampling with a random start method, with probability proportional to size. The number of households in the village will be used as the measure of size. These data are based on the population database compiled in the National Institute of Statistics, Ministry of Planning, from several sources including a gazetteer of the Geographic Department, a village file constructed in 1993 by the United Nations Transitional Authority in Cambodia (UNTAC), a population statistics of Battambang province constructed by the United Nations High Commissioner for Refugees (UNHCR), and supplemental population estimates supplied by the Ministry of Interior and the Municipality of Phnom Penh. The merger of these multiple sources constitutes the sample frame for the Socio-Economic Survey of Cambodia (SESC).

For the second stage of selection, every sample PSU is to be listed in order to provide a current frame from which to select the particular households to be interviewed. Very large PSU's or those with household size of more than 300 will be subdivided into two or more segments with at most 300 households per segment where one segment will be selected at random for listing. Upon completion of the listing of households for every PSU, a fixed sample size of exactly 10 households will be selected using a circular systematic selection with a random start from Banteay Meanchey, Pursat, Siem Reap, Phnom Penh, Other Urban and Other Rural; while 20 households will be selected from Battambang, Kampong Thom, Ratanak Kiri and Svay Rieng.

The sample will be large enough to provide reliable data for the considered ten (10) strata and for Cambodia as a whole.

## **2. Sample Selection**

### **2.1 First Stage Selection**

The first stage of sample selection will involve the drawing of sample villages from each stratum. Within each stratum, villages will be arranged by geographic codes and the number of households for every village based on the sample frame records will be cumulated. Sample villages will then be selected using the systematic sampling with random start method, with probability proportional to size (pps). The number of households in the village will be used as the measure of size. Sample village selection will be done through the use of a computer program.

The selection probability for village  $i$  in stratum  $h$  will be computed as:

$$P_{hi} = (a_h / \sum_i M_{hi}) \times M_{hi} \quad (\text{Eq. 1})$$

where

$P_{hi}$  = probability of selecting the  $i^{\text{th}}$

PSU in h stratum

$a_h$  = number of sample villages to be drawn from h stratum

$M_{hi}$  = number of households in village i as recorded in the sample frame

$M_{hi}$  = total number of households in stratum h as recorded in the sample frame  
=  $M_h$

## **2.2 Second Stage Selection**

For each sample village (PSU), a field listing operation will be undertaken except for large villages. Large villages will be segmented first, comprising about 300 households based on the current household estimates by the commune or village leaders. A segment will then be chosen randomly in which a complete listing of households will be done. This will entail carrying out a complete canvass of the PSU in order to make a current and complete listing of households contained within. The procedure will involve creating a sketch map for the PSU where physical boundaries in the village and the location of each household will be sketched. Canvassing, on the other hand, will entail a systematic covering of the entire village following a prescribed path of travel in order to make sure that all housing units in which the households reside will be accounted for.

After the listing operation is completed, a fixed sample size of 10 households will be selected in each PSU for the following stratum: Banteay Meanchey, Phnom Penh, Pursat, Siem Reap, Other Urban and Other Rural while 20 households will be selected from each PSU in Battambang, Kampong Thom, Ratanak Kiri and Svay Rieng. The selection will be carried out using a circular systematic random sampling with a random start. The sampling interval will be equal to the current household estimates in the PSU divided by 10 or 20, as the case maybe.

The selection probability for households j in village i will be computed as:

$$P_{ij} = (N_{hi} / M_{hi}^*) \quad (\text{Eq. 2})$$

where

$P_{ij}$  = probability of selecting the  $j^{\text{th}}$  households in village  $i$

$N_{hi}$  = number of households to be selected in village  $i$

= 10 households for Banteay Meanchey, Phnom Penh, Pursat, Siem Reap, Other Urban and Other Rural

= 20 households for Battambang, Kampong Thom, Ratanak Kiri, and Svay Rieng

$M_{hi}^*$  = number of households actually residing in village  $i$  in stratum  $h$  at the time of the survey. This will be equal to the number of households listed in the village  $i$  if the village is not segmented. If only a segment of village  $i$  will be listed, this is equal to the number of households listed in the selected segment.

The overall selection probability ( $f_h$ ) will be computed as follows:

$$f_h = P_{hi} \times P_{ij}$$

$$f_h = ((a_h \times M_{hi}) / M_h) \times (N_{hi} / M_{hi}^*) \quad (\text{Eq. 3})$$

$$f_h = ((a_h \times N_{hi}) / M_h) \times (M_{hi} / M_{hi}^*)$$

where

$M_h$ ,  $M_{hi}$  &  $a_h$  are as defined in Eq. 1

$N_{hi}$  and  $M_{hi}^*$  are as defined in Eq. 2

Note that if there were no difference in the number of households between the frame and at the time of the survey in any PSU, that is  $M_{hi} = M_{hi}^*$ , then  $f_h$  would be as follows:

- i) For Banteay Meanchey, Phnom Penh, Pursat, Siem Reap, Other Urban and Other Rural

$$f_h = (a_h \times 10) / M_h \quad (\text{Eq. 4})$$

- ii) For Battambang, Kampong Thom, Ratanak Kiri and Svay Rieng

$$f_h = (a_h \times 20) / M_h \quad (\text{Eq. 5})$$

where

$a_h$  = number of village/PSU  
to be drawn in stratum h

$M_h$  = total number of households  
in stratum h as recorded in  
the sample frame

### **2.3 Basic Weights**

The basic weight (that is, the weight before any other adjustments are made such as those which might be necessary for non-response), for all households in a given PSU/village, is equal to the reciprocal of the overall probability of selection,  $f_h$ , or

$$W_i = 1 / f_h$$

$$W_i = (M_h / (a_h \times N_{hi})) \times (M_{hi}^* / M_{hi}) \quad (\text{Eq. 6})$$

where

$M_h$ ,  $M_{hi}$  &  $a_h$  are as defined in Eq. 1

$N_{hi}$  and  $M_{hi}^*$  are as defined in Eq. 2

Thus, there will be a different weight for every PSU, that is, 750 separate weights for the two rounds. Likewise, the data file must provide therefore for storing values of  $M_{hi}$  and  $M_{hi}^*$  for every PSU, so that their ratio can be calculated easily.

In applying the weights to the data files in order to make estimates, every sample household and every sample person in a given PSU will receive the same weight, that is, whatever the calculated basic weight is for that PSU.

## 2.4 Weight Adjustments for Large PSU's

Since large villages or those with an estimated number of households with 300 or more will be segmented, and only one randomly selected segment will be listed, and from the selected segment, the sample of 10 or 20 households will then be selected, thus, the weight for that PSU will be as follows:

$$W_i = ((M_h \times M_{hi}^*) / (a_h \times N_{hi})) \times (M_{hi} / d_i)$$

$$W_i = (M_h / (a_h \times N_{hi})) \times ((d_i \times M_{hi}^*) / M_{hi}) \quad (\text{Eq. 7})$$

where

$d_i$  = the number of subdivisions/  
segment in village i

$M_h, M_{hi}^*, M_{hi}, a_h, N_{hi}$  are as defined  
earlier

## 2.5 Sample Size

The sampling strategy will require the selection of a total of 9,000 sample households from 750 sample villages for the two rounds. The number of PSU's and households for each stratum to be sampled are given in Table 1.

Table 1. Distribution of Sample Villages and Households by Stratum  
by Survey Round : 1996

STRATUM	FIRST ROUND		SECOND ROUND		TOTAL	
	Village	H'hold	Village	H'hold	Village	H'hold
ALL STRATA	375	4500	375	4500	750	9000
Banteay Meanchey	30	300	30	300	60	600
Battambang	20	400	20	400	40	800
Kampong Thom	20	400	20	400	40	800
Phnom Penh	60	600	60	600	120	1200
Pursat	30	300	30	300	60	600
Ratanak Kiri	15	300	15	300	60	600
Siem Reap	30	300	30	300	60	600
Svay Rieng	20	400	20	400	40	800
Other Urban	50	500	50	500	100	1000
Other Rural	100	1000	100	1000	200	2000

### 3. Estimation Procedure

#### 3.1 Estimation Procedure for Household Information

The estimate of the stratum total is given by the following formula:

$$\hat{Y}_h = \sum_{i=1}^{a_h} \sum_{j=1}^{n_{hi}} w_{hij} y_{hij} \quad \text{for } \begin{matrix} j = 1, 2, \dots, n_{hi} \\ i = 1, 2, \dots, a_h \end{matrix} \quad (\text{Eq. 8})$$

where

$\hat{Y}_h$  = estimate of characteristic **y** for stratum **h**

$y_{hij}$  = any characteristic of household **j** in sample village **i** in stratum **h**

$n_{hi}$  = number of sample households in village **i**

$a_h$  = number of sample villages in stratum **h**

$w_{hij} = 1 / f_h$

The estimate for the total for all strata, ( $\hat{Y}$ ), was computed as the sum of the estimates for each stratum. That is:

$$\hat{Y} = \sum_h \hat{Y}_h \quad h = 1 \text{ to } 10 \quad (\text{Eq. 9})$$

The estimated stratum mean is a ratio computed as:

$$r_h = \frac{\hat{Y}_h}{\hat{N}_h} = \frac{\sum_{i,j} w_{hij} y_{hij}}{\sum_{i,j} w_{hij} n_{hij}} \quad (\text{Eq. 10})$$

where

$y_{hij}$ ,  $a_h$ ,  $n_{hi}$ ,  $w_{hij}$  is as defined in Eq. 8

$$\sum_{i,j} n_{hij} = 1 \quad \text{for } j = 1, 2, \dots, n_{hi} \\ i = 1, 2, \dots, a_h$$

The population mean is also a ratio,  $r$ , which was estimated using the following formula:

$$r = \frac{\sum_{h,i,j} w_{hij} y_{hij}}{\sum_{h,i,j} w_{hij} n_{hij}} \quad (\text{Eq. 11})$$

where

$y_{hij}$ ,  $a_h$ ,  $n_{hi}$ ,  $w_{hij}$  is as defined in Eq. 8

$$\sum_{i,j} n_{hij} = 1 \quad \text{for } j = 1, 2, \dots, n_{hi} \\ i = 1, 2, \dots, a_h \\ h = 1, 2, \dots, 10$$

### 3.2. Estimation Procedure for Sector and National Level Estimates

As mentioned, only 87.26 percent of Cambodian villages were covered because of security reasons. Due to this reason, there is no way of providing

estimates for Cambodia as a whole or sectoral or regional estimates through this household sample survey.

Despite these issues which prevent the estimation of parameters from the sample data for the whole of Cambodia based on accepted statistical methodology, many users desire even approximate national level estimates. Recognizing this need, it was decided that the sample data be used to prepare estimates for areas excluded in the sample frame.

### 3.2.1 For Excluded Villages in Domain h

In the assumption that there is a parallel or equi-proportionate changes in the number of households in the two groups (included and excluded villages), the estimate of the total for the excluded rural villages for the following domains: Banteay Meanchey, Battambang, Kampong Thom, Pursat, Siem Reap and Svay Rieng will be derived using the following formula:

$$\begin{aligned} \hat{Y}_{2h} &= (M_h^* / M_h) * B_h * \bar{Y}_h && \text{(Eq. 12)} \\ &= (B_h^*) \bar{Y}_h \end{aligned}$$

where

$M_h^*$  = estimated number of households of the included rural villages in domain h based on the SESC

$M_h$  = number of households of the included rural villages in domain h in the sample frame

$B_h$  = total number of households of the excluded rural villages in domain h

$B_h^* = (M_h^* / M_h) (B_h)$

—

$$Y_h = \frac{\sum_{i,j} w_{hij} y_{hij}}{\sum_{i,j} w_{hij}}$$

= weighted mean for characteristic y in the domain h that are included in the frame

$n_{hij}$  = number of households selected in village i of domain h

$y_{hij}, a_h, w_{hij}$  is as defined in Eq. 8

For Other Urban and Other Rural domains, the estimated totals for excluded villages will be derived using the following formula:

$$\begin{aligned} \hat{Y}_{3h} &= (M_h^* / M_h) (B_h) (\hat{\bar{Y}}_h^*) && \text{(Eq. 13)} \\ &= (B_h^*) \hat{\bar{Y}}_h^* \end{aligned}$$

where

$$\hat{\bar{Y}}_h^* = \frac{(M_h^*)(\bar{Y}_h)}{(M_h^* + B_h^*)} + \frac{(B_h^*)(\bar{Y}_h')}{(M_h^* + B_h^*)}$$

$M_h^*, M_h, B_h^*$  as defined in Eq. 12

$\bar{Y}_h$  = mean for characteristic y for villages in domain h that are included in the sample frame.

$$= \frac{\sum_{i,j} w_{hij} y_{hij}}{\sum_{i,j} w_{hij}} = \bar{y}_h$$

$$\bar{Y}_h' = \sum_k (B_{hk} / B_h) \times \bar{y}_{hk}$$

= weighted mean for characteristic y for villages in domain h that are excluded in the sample frame.

$B_{hk}$  = number of households in province k of the excluded villages in domain h

$B_h$  as defined in Eq. 12

—

$$y_{hk} = \frac{1}{10} \sum_{i,j} y_{hkij} / (10 \times a_{hk})$$

= mean of province k in domain h in the excluded urban / rural villages

10 = the number of sample households selected per village in Other Urban and Other Rural domains

$a_{hk}$  = the number of selected villages in province k of domain h

The estimated total for characteristic y in the excluded urban and rural villages in domain h will be estimated as:

$$\hat{Y}_2 = \sum_{h=1}^8 \hat{Y}_{2h} + \sum_{h=9}^{10} \hat{Y}_{3h}$$

### 3.2.2 For Excluded Villages in Totally Excluded Provinces

Using the assumption that the urban/rural villages of totally excluded provinces behave the same way as in the included villages of Other Urban and Rural domains, the estimated total for characteristic y in the excluded urban or rural villages in the totally excluded provinces will be derived using the following formula:

$$\hat{Y}_{4h} = (M_h^* / M_h) A_h \times \frac{\hat{Y}_h}{M_h^*}$$

where

$M_h^*$ ,  $M_h$  as defined in Eq. 12 (based on the Other Urban/Other Rural  $M_h^*$  and  $M_h$ )

$A_h$  is the number of households in the urban/rural villages in totally excluded provinces based on the UNTAC population data file.

$\frac{\hat{Y}_h}{M_h^*}$  is as defined in Eq. 13

The estimated total for characteristic y in the excluded urban/rural villages in the totally excluded provinces in the sample frame will be computed as:

$$\hat{Y}_3 = \sum_h \hat{Y}_{4h}$$

### 3.2.4 Extrapolated Estimate for Cambodia

The estimates for the whole Cambodia will be arrived at by adding the estimates based on the sample frame and the extrapolated data for the villages excluded in the frame as follows:

$$\hat{Y} = \hat{Y} + \hat{Y}_2 + \hat{Y}_3$$

where

$\hat{Y}$  = estimate for villages included in the sample frame

$\hat{Y}_2$  = extrapolated estimate for excluded villages in the domain

h

$\hat{Y}_3$  = extrapolated estimate for excluded villages in the totally excluded provinces

### 3.3 Estimation of Variances

In order to provide a basis for assessing the reliability or precision of SESC survey estimates, the estimation of the magnitude of sampling error in the survey data shall be computed. Since most of the estimates from the survey are in the form of weighted ratios, thus variances for ratio estimates will thus be presented. The procedures in deriving the estimates for the variances are described below.

All variances of the ratio estimates will be given of the form:

$$\text{var}(r) = \frac{1}{X^2} (1 - f_h) \left( \frac{a_h}{a_h - 1} \right) (z_{hi}^2 - z_h^2/a_h) \quad (\text{Eq. 14})$$

where

$$r = \frac{\sum_j w_{hij} y_{hij}}{\sum_j w_{hij} x_{hij}}$$

$$X^2 = \left( \sum_{hij} w_{hij} x_{hij} \right)^2$$

$$z_{hi} = y_{hi} - r x_{hi}$$

$$y_{hi} = \sum_j w_{hij} y_{hij}$$

$$x_{hi} = \sum_j w_{hij} x_{hij}$$

$$a_h = \text{number of sample villages from stratum } h$$

$$w_{hij} = \text{weight for each individual in the sample household}$$

The variance of the ratio estimate,  $r_h$ , in stratum  $h$  is given by the formula:

$$\text{var}(r_h) = \frac{1}{x_h^2} (1 - f_h) \left[ \frac{a_h}{a_h - 1} \right] \sum_i (z_{hi}^2 - z_h^2/a_h) \quad (\text{Eq. 15})$$

where

$$x_h = \sum_j w_{hij} x_{hij}$$

$f_h, a_h, z_{hi},$  and  $z_h$  are as defined earlier

### **3.4 Standard Errors and Coefficient of Variation**

For analytical and evaluation purposes, the standard error (SE) and relative standard error or coefficient of variation (CV) are commonly used measures of sampling error. Both measures are derived from the estimates of variances described above.

The standard error of a survey estimate provides a measure of how far the survey estimate is likely to vary from the true population value (i.e., parameter) as a result of having collected the data on a sample basis rather than through a complete census.

Computationally, the standard error of a survey estimate is obtained by taking the square root of the estimated variance of the survey estimate, or

$$se(r) = \text{var}(r)^{1/2} \quad (\text{Eq. 16})$$

The relative standard error or coefficient of variation (CV), on the other hand, provides a measure of the relative variance of a survey estimate; that is, the magnitude of the estimated sampling error relative to the magnitude of the estimate itself. Its primary function is to enable the user of the data to compare the relative reliability or precision with which different types of survey characteristics have been measured (e.g. means versus proportions). In such cases, the direct comparison of standard errors tends to be uninformative since the magnitude of the standard error is dependent upon the magnitude of the estimate. The CV, which is usually expressed as a proportional error, provides the means of overcoming this non-comparability problem.

Computationally, the coefficient of variation is calculated as:

$$CV(R) = se(r) / r \quad (\text{Eq. 17})$$

### **3.5 Valid Use of the Ratio Estimate**

A reasonable control of the sample size is needed for the valid use of the ratio mean. This means that a control on the coefficient of variation of the sample size is required.

In using ratio estimation, the coefficient of variation (CV) of the denominator should not exceed 0.2. A CV that is < 0.1 is sufficiently good.

The coefficient of variation of x is given by:

$$CV(x) = se(x) / x \quad (\text{Eq. 18})$$

$$= 1/x \{ (1 - f_h) [ a_h / (a_h - 1) ] (x_{hi}^2 - x_h^2/a_h) \}^{1/2}$$

where

$f_h$ ,  $a_h$ ,  $x_{hi}$ , and  $x_h$  are as defined earlier

### 3.6 Design Effect (DEFT)

To measure the efficiency of the design adopted for SESC, the design effect (DEFT) shall be computed. The design effect is the ratio of the variance based on the adopted design (var r) and the variance based on a simple random sampling (var  $r_s$ ). A DEFT value of 1.0 indicates that the sample design adopted is as efficient as a simple random sample while a value greater than 1.0 indicates the increase in the sampling error due to the use of a less statistically efficient design. More so, a DEFT value of less than 1.0 indicates that the sample design used is more efficient than a simple random sample.

The variance of r based on equivalent simple random sample is of the form:

$$\text{var}(r_s) = [1 / (n - 1)] \sum (w_{hij} * z_{hij}^2) / w_{hij} \quad (\text{Eq. 19})$$

$$\text{where } z_{hij} = y_{hij} - r * x_{hij}$$

n = number of samples

is taken over all elements in the sample

Computationally, the design effect (DEFT) is calculated as:

$$\text{DEFT} = \text{var}(r) / \text{var}(r_s) \quad (\text{Eq. 20})$$

## 4. **Software That Can Be Used In the Computation of Sampling Errors**

The following softwares have the facility to compute variances taking into account the complex design of the survey:

- a) **CLUSTERS** developed by Vijay Verma and Mick Pearce for the World Fertility Survey. This software has already been used in the 1993-94 SESC.
- b) **SUDAAN** from Research Triangle Institute (RTI)
- c) **CENVAR** from the US Bureau of Census
- d) **OSIRIS** from the University of Michigan

## 5. Computational Procedure

A tailored program may also be prepared following the procedure outlined below:

### 5.1 Computation of variance for ratio $r_h$ in stratum $h$ taking into consideration the design of the SESC

#### 5.1.1 Preliminary estimates/data required at the stratum level

- a) Number of sample villages for each stratum ( $a_h$ )
- b) Sampling fraction ( $f_h$ )
- c) Stratum estimates for the total variables involved in the computation of ratios, means or proportions

$$\hat{Y}_h = \mathbf{y}_h = \sum W_{hij} * y_{hij}$$

$$\hat{X}_h = \mathbf{x}_h = \sum W_{hij} * x_{hij}$$

- d) Ratio ( $r_h$ )

$$r_h = \left( \sum W_{hij} * y_{hij} \right) / \left( \sum W_{hij} * x_{hij} \right)$$

$$= \mathbf{y}_h / \mathbf{x}_h$$

**5.1.2 Estimates to be computed for each sample village (i) in stratum h**

- a. weighted total for characteristics y ( $y_{hi}$ ) and characteristics x ( $x_{hi}$ )

$$y_{hi} = \sum_j W_{hij} * y_{hij}$$

$$x_{hi} = \sum_j W_{hij} * x_{hij}$$

- b.  $z_{hi} = y_{hi} - r_h * x_{hi}$

**5.1.3 Last Steps for the Variance Computation for  $r_h$**

- a. Add all  $z_{hi}$  in stratum h

$$z_{hi} = z_h$$

- b. Compute  $D_{hi} = z_{hi}^2 - z_h^2/a_h$

$$a_h = \text{number of sample villages in stratum h}$$

- c. Generate  $D_h = \sum_i D_i$

- d. For each stratum, compute:

$$V_h = (1 - f_h) [a_h / a_h - 1] D_h$$

- e. The variance of  $r_h$  is given by:

$$\text{Var} (r_h) = [1 / x^2] v_h$$

**5.2 *Computational Procedure for the Variance of the Ratio r for All Areas in the Sampling Frame***

The variance of the ratio r, where

$$r = \left( \sum_{h i j} W_{hij} * y_{hij} \right) / \left( \sum_{h i j} W_{hij} * x_{hij} \right)$$

can be computed using the procedure outlined for the computation of the variance of the ratio  $r_h$  with the following modifications:

a) Use  $r$  instead of  $r_h$  in the formula

b) Use  $\hat{X}$  instead of  $X_h$  or  $\hat{Y}$  instead of  $Y_h$  where

$$\hat{Y} = \hat{y} = \sum_{h=1}^H \sum_{i=1}^{a_h} \sum_{j=1}^{D_i} W_{hij} * y_{hij}$$

$$\hat{X} = \hat{x} = \sum_{h=1}^H \sum_{i=1}^{a_h} \sum_{j=1}^{D_i} W_{hij} * x_{hij}$$

For each stratum, compute

$$V_h = [a_h / (a_h - 1)] D_i$$

$$= [a_h / (a_h - 1)] (z_{hi}^2 - z_h^2 / a_h)$$

$$d. \text{Var} [r] = 1 / x^2 \sum_{h=1}^H V_h$$

An alternative computational procedure would be to express the  $z_h$  and  $z_{hi}$  in terms of  $x_{hij}$  and  $y_{hij}$  in the general formula for the variances instead of computing  $z_h$  and  $z_{hi}$  separately.

### 5.3 Computation of variance, $\text{var}(r_s)$ , for an equivalent simple random sample

$$\text{var}(r_s) = [1 / (n - 1)] \sum_{h=1}^H (W_{hij} * z_{hij}^2) / W_{hij}$$

a. For each sample household, compute

$$(1) z_{hij} = y_{hij} - r * x_{hij}$$

where

$y_{hij}$  = characteristic  $y$  for household  $j$  in village  $i$  and stratum  $h$

$x_{hij}$  = characteristic  $x$  for household  $j$  in village  $i$  and stratum  $h$

$r$  = ratio estimate

$$= y_{hij} / x_{hij}$$

(The same result can be obtained if computation is done for each individual in the sample household)

$$(2) Q = w_{hij} * z_{hij}^2$$

b. Compute var ( $r_s$ )

$$\begin{aligned} \text{var}(r_s) &= [1 / (n-1)] \sum w_{hij} Q / \sum w_{hij} \\ &= [1 / (n - 1)] \sum (w_{hij} * z_{hij}^2) / \sum w_{hij} \end{aligned}$$

#### 5.4 Compute DEFT

The design effect (DEFT) for estimates in all areas included in the sampling frame is given by:

$$\text{DEFT} = \text{var}(r) / \text{var}(r_s)$$

#### References:

- Kish, Leslie. Survey Sampling. New York, John Wiley and Sons, 1965.
- Luis, Elma. Final Report, Computation of Standard Errors for the Socio-Economic Survey of Cambodia, 1993-94, Cambodia. 1993
- Morada, Marietta. Final Report, Sampling Design for the Socio-Economic Survey of Cambodia - 1993/94, Cambodia. 1993.
- Verma, Vijay and Pearce, Pearce. CLUSTERS, A Package Program for the Computation of Sampling Errors for Clustered Samples, Version 3, User's Manual. February 1993.