Estimating Sampling Errors

of Means, Total Fertility, and Childhood Mortality Rates Using SAS

MEASURE *DHS+*

ESTIMATING SAMPLING ERRORS OF MEANS, TOTAL FERTILITY, AND CHILDHOOD MORTALITY RATES USING SAS

Mamadou Thiam and Alfredo Aliaga

ORC Macro
Calverton, Maryland

June 2001

Acknowledgments

The authors wish to thank Gora Mboup, Noureddine Abderrahim, and Shea Rutstein for their valuable contribution. We also wish to thank Monique Barrère for the French translation.

1 Introduction

The Demographic and Health Surveys (DHS) program conducts surveys in developing countries to provide information mainly on family planning and maternal and child health. DHS data, like many survey data, are collected based on complex probability sample designs featuring disproportionate sample allocation, stratification, clustering, and the selection of different types of sampling units. Survey estimates are subject to two types of errors: nonsampling errors and sampling errors. Nonsampling errors, which are difficult to measure, generally result from mistakes made during data collection and data processing. Sampling errors, which we deal with in this report, arise from limiting the inquiry to a sample of the population. If a probability sample is drawn, the particular sample obtained is only one of the many samples that could have been selected from the same population using the same sample design. Each of these samples would yield estimates that differ somewhat from the estimates derived from the sample that is actually available. The sampling error of an estimate is a measure of the variability between estimates from all possible samples of the same size and design under the same essential survey conditions. The degree of this variability can be estimated from the one probability sample that is selected. The sample design features influence the extent of sampling errors and must be taken into account in estimating sampling errors.

This technical report presents three user-friendly SAS macros for the computation of sampling error estimates: MEAN, TFR, and MORT. The MEAN macro computes the sampling error of means and proportions while TFR and MORT estimate those of fertility and childhood mortality rates, respectively. All three macros produce the sampling error for the total sample as well as for specified subclasses. In addition to estimating the sampling error, the macros also produce the unweighted and weighted sample sizes, the sampling error under simple random sampling, the design effect (except for the total fertility rate), the relative error and the 95% confidence intervals of the estimate. Large-scale surveys produce a large number of estimates, but sampling errors can be computed and presented in the survey report only for selected variables. The MEAN macro will allow analysts to easily estimate the sampling error of means and proportions for any variable of interest in the context of complex survey design. The TFR and MORT macros provide SAS users with an additional tool, which is not yet available in the current version of SAS. Each macro is described in detail in the following sections.

2 The MEAN macro

2.1 Computational method and formulae

The MEAN macro uses the Taylor linearization method of variance estimation to estimate sampling errors of means and proportions under an ultimate cluster sample selection model. Under the ultimate cluster sampling model, elements within primary sampling units are divided into ultimate clusters of roughly equal size, and then a sample of ultimate clusters is taken without replacement from every PSU. The Taylor linearization method treats any estimator as a ratio estimator, obtains a linear approximation of the estimator, and then uses the variance of the approximation to estimate the variance of the estimator itself. This method also assumes that two or more clusters are selected independently and with replacement from each stratum. In practice, clusters are usually selected without replacement, and variance estimates obtained using the Taylor linearization method are then overstated, but this should be negligible if the sampling fraction is small. With the Taylor linearization method, aggregated quantities are computed only at the cluster level; the design information at other subsequent stages is not used. It is worth noting that the strata used for computing sampling errors are not necessarily the explicit strata used in sample design and selection. Strata to be used for the computation of sampling errors may be created by grouping clusters based on some characteristic of similarity so as to maintain homogeneity within each

stratum. In DHS surveys, each stratum is formed by grouping two or three clusters, thus further refining the original stratification used in the sample design.

To supply the basic formulae, we consider a ratio estimator r = y/x where y and x are weighted sums over the whole sample or a subclass. Suppose there are H strata in the sample or subclass, and m_h clusters are selected from stratum h. For any cluster i in stratum h, the notations follow:

y_{hij} = value of variable y for element j in cluster i in stratum h;

y_{hi} = weighted sum of all values y_{hij} for all elements in cluster i, i.e.,

$$y_{hi} = \sum_{j=1}^{n_{hi}} w_{hij} y_{hij}$$

where w_{hij} is the sample weight for element j in cluster i in stratum h, and n_{hi} is the number of elements in cluster i.

y_h = sum of the values of y_{hi} for stratum h, i.e.,

$$y_h = \sum_{i=1}^{m_h} y_{hi}$$

y = sum over the whole sample or subclass, i.e.,

$$y = \sum_{h=1}^{H} y_h$$

Similar notations apply for variable x.

The variance of the ratio estimator r is computed using the formula given below, with the sampling error being the square root of the variance:

$$SE^2(r) = \text{var}(r) = \frac{1}{x^2} \sum_{h=1}^{H} \left[\frac{(1-f_h)m_h}{m_h - 1} \left(\sum_{i=1}^{m_h} z_{hi}^2 - \frac{z_h^2}{m_h} \right) \right] \qquad (1)$$

in which $z_{hi} = y_{hi} - r.x_{hi}$ and $z_h = y_h - r.x_h$

where f_h is the stratum sampling fraction, which is so small, particularly for large sample sizes that it can be ignored. Note that x_{hij} is equal to 1 in the MEAN macro.

The estimate of the sampling error given by the Taylor linearization method is not unbiased. The magnitude of this bias depends on the coefficient of variation (CV) of cluster sizes and may be ignored for variation less than 0.2. The MEAN macro prints the coefficient of variation of cluster sizes for the entire sample as well as for every subclass. This CV is computed using the following formula:

$$CV = \frac{1}{x}\left[\sum_{h=1}^{H}\frac{(1-f_h)m_h}{m_h-1}\left(\sum_{i=1}^{m_{hi}}x_{hi}^2-\frac{(\sum_{i=1}^{m_{hi}}x_{hi})^2}{m_h}\right)\right]^{\frac{1}{2}} \qquad (2)$$

where x_{hi} is the sum of the weights for all elements in cluster i and stratum h.

The design effect (DEFT) printed by the MEAN macro is defined in DHS surveys as the ratio of the estimated sampling error under the actual sample design to the sampling error that would be obtained under simple random sampling. Assuming simple random sampling, the sampling error for the ratio estimator r = y/x is given by the following formula:

$$SE_{srs}^2(r) = \frac{1}{n-1}\frac{\sum_{h=1}^{H}(1-f_h)\sum_{i=1}^{m_h}\sum_{j=1}^{n_{hi}}w_{hij}z_{hij}^2}{\sum_{h=1}^{H}\sum_{i=1}^{m_h}\sum_{j=1}^{n_{hi}}w_{hij}} \qquad (3)$$

where $z_{hij} = y_{hij} - r.x_{hij}$

2.2 Syntax and input

The following statement, in which key parameters are to be specified, will invoke and execute the MEAN macro within a SAS session:

%MEAN (DATA=, WEIGHT=, PSU=, STRATA=, BYVAR=, ALLVARS=);

In the above macro call, the *DATA=* statement names the input data set, which should contain all variables for which the sampling error is requested, and any other variables identified in the *WEIGHT, PSU, STRATA,* and *BYVAR* statements. The *WEIGHT=* statement identifies the weight variable. The *PSU=* statement specifies the primary sampling unit to which each ultimate cluster in the sample belongs, whereas *STRATA=* names the stratification variable to be used for estimating sampling errors. If sampling errors are to be estimated for subclasses, the *BYVAR=* statement must name the variable defining the subclasses. All variables for which sampling errors are to be computed should be listed and separated with a space in the *ALLVARS=* statement. It should be noted that except the *BYVAR=* statement which is optional, all other statements are required for the macro to work properly. If subclass estimates are requested, then the macro will produce estimates for all subclasses as well as for the whole sample.

2.3 Output

The MEAN macro creates and prints an SAS data set containing the following results for every variable for which the sampling error was requested:

CV : Coefficient of variation of cluster sizes.

VARIABLE : Variable name.

LABEL : Variable label, if any. The maximum length is 40 characters including spaces.

TYPE : Type of statistics (mean or proportion).

MEAN : Weighted sample mean or proportion.

STDERROR : Sampling error of the sample mean or proportion.

N_UNWGT : Unweighted number of cases used in the calculations.

N_WGT : Weighted number of cases used in the calculations.

SRS : Sampling error of the mean or proportion under simple random sampling.

DEFT : Design effect (ratio of sampling errors).

RELERROR : Relative error of the sample mean or proportion.

LOWER : Lower bound of the 95% confidence intervals.

UPPER : Upper bound of the 95% confidence intervals.

The resulting estimates were compared with estimates obtained through the Taylor linearization method used in SUDAAN and the sampling error module of the Integrated System for Survey Analysis (ISSA). All three programs produced the same estimates.

Examples

The examples in this report use the data set from the 1997 Madagascar Demographic and Health Survey. The survey objectives include the estimation of the prevalence of contraceptive use in women 15-49 years old and the estimation of fertility and childhood mortality rates. For the purpose of the illustration, the following selected variables are used:

V102 : Type of residence (1 = urban, 2 = rural)
V106 : Highest educational level attended (0 = no education, 1 = primary, 2 = secondary , 3 = higher)
V201 : Total number of children ever born alive to the woman
V502 : Marital status (0 = never married, 1 = currently married, 2 = formerly married)
V613 : Ideal number of children
V005 : Sample weight.

A sample of 7,060 women with completed interviews was obtained using a stratified 2-stage design. In the first stage, 270 enumeration areas (EAs) were selected. A complete listing of households within selected EAs was carried out. The lists of households obtained served as the sampling frame for the selection of households in the second stage. All eligible women identified in the selected households were included in the sample.

Categorical variables need to be recoded in order to estimate the proportions that are of interest to the survey. The following statements create the working data set TEST that will contain the new variables to be used in the macro call. It should be noted that in the original data set (MDIR), the value of the sample weight is multiplied by 1,000,000.

```
                    /***** RECODING VARIABLES ******/

DATA TEST; SET MDIR;
        RWEIGHT=V005/1000000;           /* weight variable*/
        EVBORN = V201;
        IF V102=1 THEN URBAN=1; ELSE URBAN=0;
        IF V106 IN (2,3) THEN EDUC=1; ELSE EDUC=0;
        IF 0<=V613<=45 THEN IDEAL=V613; ELSE IDEAL=.;
        IF V502=1 THEN CURMAR=1; ELSE CURMAR=0;
        IF 1<=V313 <=3 THEN CUSE=1 ; ELSE CUSE=0 ;
        LABEL URBAN     = 'Urban residence'
              EDUC      = 'With secondary education or higher'
              CURMAR    = 'Currently married (in union)'
              CUSE      = 'Currently using any method'
              EVBORN    = 'Children ever born'
              IDEAL     = 'Ideal number of children';
RUN;
```

Example 1: Sampling errors for subclasses are not requested.

```
LIBNAME in 'J:\USER\';
OPTIONS MSTORED SASMSTORE= in;
%MEAN (DATA= test, WEIGHT= rweight, PSU=v021, STRATA=v022,
        ALLVARS= urban educ cuse evborn ideal);
```

In this example, the LIBNAME and OPTIONS statements bring into the SAS session the MEAN macro, which is stored in the directory 'J:\USER'. Sampling errors will be produced when the macro call statement %MEAN is executed.

Example 2: Sampling errors for subclasses are requested.

```
LIBNAME in 'J:\USER\';
OPTIONS MSTORED SASMSTORE= in;

%MEAN (DATA=test, WEIGHT= rweight, PSU=v021, STRATA=v022,
        BYVAR=v102, ALLVARS=urban educ cuse evborn ideal);
```

The output files for these two examples are presented below.

Output 1: Sampling errors for subclasses are not requested

COEFFICIENT OF VARIATION FOR CLUSTER SIZES: ENTIRE SAMPLE

VARIABLE	CV
CUSE	0.030
EDUC	0.029
EVBORN	0.029
IDEAL	0.029
URBAN	0.029

SAMPLING ERRORS: ENTIRE SAMPLE

VARIABLE	LABEL	TYPE	MEAN	STDERROR	N_UNWGT	N_WGT	SRS	DEFT	RELERROR	LOWER	UPPER
CUSE	Currently using any method	Proportion	0.628	0.008	7060	7060	0.006	1.328	0.012	0.613	0.643
EDUC	With secondary education or higher	Proportion	0.269	0.012	7060	7060	0.005	2.310	0.045	0.244	0.293
EVBORN	Children ever born	Mean	3.215	0.050	7060	7060	0.038	1.308	0.016	3.115	3.315
IDEAL	Ideal number of children	Mean	5.306	0.082	6447	6358	0.035	2.303	0.015	5.142	5.469
URBAN	Urban residence	Proportion	0.281	0.011	7060	7060	0.005	2.007	0.038	0.260	0.303

Output 2: Sampling errors for subclasses are requested (urban and rural)

COEFFICIENT OF VARIATION FOR CLUSTER SIZES: ENTIRE SAMPLE

VARIABLE	CV
CUSE	0.030
EDUC	0.029
EVBORN	0.029
IDEAL	0.029
URBAN	0.029

SAMPLING ERRORS: ENTIRE SAMPLE

VARIABLE	LABEL	TYPE	MEAN	STDERROR	N_UNWGT	N_WGT	SRS	DEFT	RELERROR	LOWER	UPPER
CUSE	Currently using any method	Proportion	0.628	0.008	7060	7060	0.006	1.328	0.012	0.613	0.643
EDUC	With secondary education or higher	Proportion	0.269	0.012	7060	7060	0.005	2.310	0.045	0.244	0.293
EVBORN	Children ever born	Mean	3.215	0.050	7060	7060	0.038	1.308	0.016	3.115	3.315
IDEAL	Ideal number of children	Mean	5.306	0.082	6447	6358	0.035	2.303	0.015	5.142	5.469
URBAN	Urban residence	Proportion	0.281	0.011	7060	7060	0.005	2.007	0.038	0.260	0.303

COEFFICIENT OF VARIATION FOR SAMPLING UNIT SIZES: SUBCLASSES

----------V102=Urban ----------

VARIABLE	CV
CUSE	0.050
EDUC	0.038
EVBORN	0.038
IDEAL	0.039
URBAN	0.038

----------V102=Urban ----------

VARIABLE	CV
CUSE	0.036
EDUC	0.037
EVBORN	0.037
IDEAL	0.037
URBAN	0.037

SAMPLING ERRORS: SUBCLASSES

----------V102=Urban ----------

VARIABLE	LABEL	TYPE	MEAN	STDERROR	N_UNWGT	N_WGT	SRS	DEFT	RELERROR	LOWER	UPPER
CUSE	Currently using any method	Proportion	0.570	0.015	2376	1987	0.010	1.450	0.026	0.540	0.599
EDUC	With secondary education or higher	Proportion	0.526	0.031	2376	1987	0.010	3.043	0.059	0.463	0.588
EVBORN	Children ever born	Mean	2.496	0.070	2376	1987	0.057	1.225	0.028	2.356	2.636
IDEAL	Ideal number of children	Mean	4.181	0.115	2255	1957	0.046	2.469	0.027	3.951	4.410
URBAN	Urban residence	Proportion	1.000	0.000	2376	1987	0.000	.	0.000	1.000	1.000

----------V102=Urban ----------

VARIABLE	LABEL	TYPE	MEAN	STDERROR	N_UNWGT	N_WGT	SRS	DEFT	RELERROR	LOWER	UPPER
CUSE	Currently using any method	Proportion	0.651	0.009	4684	5073	0.007	1.314	0.014	0.633	0.669
EDUC	With secondary education or higher	Proportion	0.168	0.010	4684	5073	0.005	1.884	0.061	0.147	0.189
EVBORN	Children ever born	Mean	3.496	0.061	4684	5073	0.048	1.260	0.017	3.374	3.619
IDEAL	Ideal number of children	Mean	5.770	0.105	4192	4501	0.046	2.306	0.018	5.560	5.980
URBAN	Urban residence	Proportion	0.000	0.000	4684	5073	0.000	.	.	0.000	0.000

2.4 Expansion of the MEAN macro

The MEAN macro can also be used to compute the sampling error of more complex proportions. Suppose it is of interest to estimate the proportion of children under three with diarrhea in the last two weeks who received a medical treatment. For this example, in order to use the MEAN macro, a two-step recoding of variables is needed as is shown in the following SAS program:

Example 3

```
Data test2; set child;

                /**** First step ****/

if b5 = 0 or age_chld >= 36 then diar2w = .;
if b5 = 1 and h11 = 2 and age_chld <= 35 then diar2w = 1;
if b5 = 1 and h11^= 2 and age_chld <= 35 then diar2w = 0;

                /**** Second step ***/

if diar2w = 1 and h12z = 1 then medtre = 1;
if diar2w = 1 and h12z ^= 1 then medtre = 0;

label diar2w   = 'Had diarrhea in the last 2 weeks'
      age_chld = 'Child age'
      b5       = 'Child alive'
      h11      = 'Child had diarrhea in the last 24 hours or within the last 2 weeks, but not the
                  last 24 hours'
      h12z     = 'Child taken to a medical facility for treatment of the diarrhea'
      medtre   = 'Sought medical treatment';
Run;
```

In the first step, variable *diar2w* is created to identify living children under 36 months with diarrhea in the 2 weeks preceding the survey. Some of these children received a medical treatment and are further identified in the second step by using variable *medtre*. When submitted, the following statements compute the sampling error:

```
LIBNAME in 'J:\USER\';
OPTIONS MSTORED SASMSTORE= in;

%MEAN (DATA=test2, WEIGHT= rweight, PSU=v021, STRATA=v022,
       BYVAR=v102, ALLVARS=medtre);
```

3 The TFR macro

3.1 Computational method and formulae

The computation of the total fertility rate (TFR) is based on the complete maternal birth histories collected in the reproduction section of the DHS questionnaire. The TFR for a period preceding the survey is calculated in DHS surveys using 5-year age groups: 15-19, 20-24, 25-29, 30-34, 35-39, 40-44, and 45-49. For the same period, the number of live births to women in each age group and the number of woman-years of exposure to childbearing in the same age group are used to calculate the total fertility rate as follows:

$$TFR(t) = 5 \times \sum_i \frac{B(i,t)}{E(i,t)}$$

where i is the age group, B(i,t) is the number of live births to women in age group i during the period t, and E(i,t) is the number of woman-years of exposure to childbearing among women in age group i during the period t. It should be noted that births that occurred during the month of interview are excluded.

In calculating B(i,t), the TFR macro creates a temporary child data file and a counting variable for the births from the input woman data file. For a particular age group i = [a, b], the counting variable is set to 1 when the birth occurs during the period of interest and the mother's age at the time of the birth is between a and b; otherwise, the counting variable is set to zero. Then B(i,t) is obtained by taking the weighted sum of the values for the counting variable as follows:

$$B(i,t) = \sum_j w_j I_{[a,b]}$$

where $I_{[a,b]} = 1$ if $a \leq$ MOM_AGE $\leq b$ and $0 <$ CHD_AGE $\leq t$
$= 0$ otherwise

and MOM_AGE is the mother's age at the time of birth, CHD_AGE is the child's age at interview, and w_j is the sample weight for the woman.

The number of woman-years of exposure to childbearing among women in age group i = [a, b] during the period t depends on the mother's age at interview (AGE_MAX) and the mother's age at the beginning of the period of interest (AGE_MIN). E(i,t) is calculated as follows:

$$E(i,t) = \sum_j w_j EXP\_ab$$

in which w_j is the sample weight for the woman, and

EXP_ab = [b – maximum(a, AGE_MIN)]/12 if AGE_MIN < b < AGE_MAX

= [AGE_MAX – maximum(a, AGE_MIN)]/12 if a < AGE_MAX < b

In surveys that included only ever-married women, the denominator of the TFR is inflated to encompass all women. The inflation factor is the proportion of ever-married women age 15-49 calculated using the household schedule.

The TFR macro uses the Jackknife repeated replication method to estimate the sampling error of the total fertility rate. The TFR is estimated from each replication of the original sample, and each replicate considers all but one cluster in the calculation of the estimate. The variability among these replicate estimates is used to compute the sampling error of the TFR as is shown in the following formula:

$$SE^2(TFR) = \text{var}(TFR) = \frac{1}{k(k-1)} \sum_{i=1}^{k} (r_i - r)^2 \qquad (4)$$

in which

$$r_i = k\,r - (k-1)\,r_{(i)}$$

and k is the total number of clusters,
r is the estimate computed from the full sample of k clusters, and
$r_{(i)}$ is the replicate estimate computed from the replicate sample of k-1 clusters (i^{th} cluster excluded).

3.2 Syntax and input

The TFR macro is invoked and executed by submitting in an SAS session the following statement in which the parameter values are to be specified:

%TFR (DATA =, PSU=, YEARS =, BYVAR=, WEIGHT =, INFLATE=);

where DATA is the name of the SAS data set, PSU identifies the cluster, YEARS is the number of years preceding the survey, BYVAR is the variable identifying the subclasses for which separate results are requested, and WEIGHT identifies the weight variable. The INFLATE= statement names the inflation factor to be used for ever-married woman samples. Except for the INFLATE= and BYVAR= statements, all other statements are required. If both the BYVAR= and INFLATE= statements are specified, the TFR macro produces results only for each level of the variable specified in the BYVAR= statement. If the BYVAR= statement is specified and the INFLATE= statement is omitted, the macro produces results for the entire sample and for each level of the variable specified in the BYVAR= statement.

In addition to the variables used in the PSU=, WEIGHT=, BYVAR=, and INFLATE= statements, the SAS data set should also contain the following variables:

CASEID : Respondent identification variable
V008 : Date of interview (CMC)
V011 : Date of birth of the respondent (CMC)
V201 : Total number of children ever born
B3 : Child's date of birth (CMC)
B5 : Survival status of the child at the time of interview (1= alive, 0=dead)
B7 : Age at death of the child (CMC), and

Any variable that will be used to identify the subclasses in the BYVAR= statement.

The TFR macro creates and prints a data set containing the total fertility rate and its sampling error, the number of weighted cases, the relative error, and the lower and upper bounds of the 95% confidence intervals. For the total fertility rate, the number of unweighted cases is not relevant because there is no known unweighted value for woman-years of exposure to childbearing.

Example 4

As in example 1, the same woman data set (*test*) is used, and the sampling errors for subclasses are not requested.

```
LIBNAME in 'J:\USER\';
OPTIONS MSTORED SASMSTORE= in;

%TFR (DATA = test, PSU = v021, YEARS = 3, WEIGHT = rweight);
```

Example 5

In this example, the sampling errors for subclasses are requested.

```
LIBNAME in 'J:\USER\';
OPTIONS MSTORED SASMSTORE= in;

%TFR (DATA = test, PSU = v021, YEARS = 3, BYVAR = v102, WEIGHT = rweight);
```

The output files are presented below.

Output 4: Sampling errors for subclasses are not requested

---------- TOTAL=Entire sample ----------

VARIABLE	LABEL	TYPE	MEAN	STDERROR	N_UNWGT	N_WGT	RELERROR	LOWER	UPPER
TRF	TOTAL FERTILITY RATE	Rate	5.969	0.140	NA	19801	0.023	5.689	6.249

Output 5: Sampling errors for subclasses are requested (urban and rural)

---------- TOTAL=Entire sample ----------

VARIABLE	LABEL	TYPE	MEAN	STDERROR	N_UNWGT	N_WGT	RELERROR	LOWER	UPPER
TRF	TOTAL FERTILITY RATE	Rate	5.969	0.140	NA	19801	0.023	5.689	6.249

----------V102=Urban ----------

VARIABLE	LABEL	TYPE	MEAN	STDERROR	N_UNWGT	N_WGT	RELERROR	LOWER	UPPER
TRF	TOTAL FERTILITY RATE	Rate	4.190	0.209	NA	5554	0.050	3.771	4.609

---------- V102=Rural ----------

VARIABLE	LABEL	TYPE	MEAN	STDERROR	N_UNWGT	N_WGT	RELERROR	LOWER	UPPER
TRF	TOTAL FERTILITY RATE	Rate	6.662	0.153	NA	14246	0.023	6.356	6.967

4 The MORT macro

4.1 Computational method and formula

The MORT macro computes the sampling error of the following five childhood mortality rates only for the five or ten years preceding the survey:

Neonatal mortality rate: the probability of dying between birth and exact age of one month

Post-neonatal mortality rate: the probability of dying between exact ages of 1 and 11 months

Infant mortality rate: the probability of dying between birth and exact age of one year

Child mortality rate: the probability of dying between exact ages of one and five years

Under-five mortality rate: the probability of dying between birth and exact age of five years.

The computation of these rates in the MORT macro follows the procedure developed by Rutstein (1984). Probabilities of death during a specific period of interest are obtained from probabilities calculated for 8 age intervals: less than 1 month, 1-2 months, 3-5 months, 6-11 months, 12-23 months, 24-35 months, 36-47 months, and 48-59 months. The probability of death for any age interval is defined as the ratio of the number of deaths occurring to children who were exposed to death in the age interval to the number of children exposed. The probability of death for a subinterval is obtained by subtracting the probability of surviving from 1. The MORT macro calculates the probability of surviving for each subinterval and the resulting probabilities are aggregated to obtain the mortality rate using the following formula:

$$(n)q(x) = 1 - \prod_{i=x}^{i=x+n}(1 - q(i)) \qquad (5)$$

where (n)q(x) is the probability of death between ages x and x + n and q(i) is the subinterval probability of death.

The MORT macro closely approximates the post-neonatal mortality rate by subtracting the neonatal mortality rate from the infant mortality rate.

The sampling error of each mortality rate is calculated for the five or ten years preceding the survey using the jackknife replication method (formula (4)). For each rate, the number of cases used in the calculation corresponds to the number of children exposed to death between the age limits during the period of interest. For both neonatal and post-neonatal rates, the number of cases is the minimum of the numbers of children exposed to death for the age intervals less than 1 month, 1-2 months, 3-5 months, and 6-11 months. For the child mortality rate, the number of cases is the minimum of the numbers of children exposed to death for the age intervals 12-23 months, 24-35 months, 36-47 months, and 48-59 months. For the under-five mortality rate, the number of cases is the minimum number of children exposed to death for all of the eight age intervals. The unweighted number of cases is used to compute the sampling error under simple random sampling for each mortality rate using the following formula:

$$SE(rate) = \left[\frac{rate(1-rate)}{n}\right]^{\frac{1}{2}} \qquad (6)$$

in which n is the unweighted number of cases.

4.2 Syntax and input

To compute the sampling errors of the childhood mortality rates, submit the following statement within an SAS session after specifying the parameter values:

%MORT (DATA =, PSU =, YEARS =, BYVAR =, WEIGHT =);

The DATA= statement identifies the name of the child data file, the PSU= statement names the sample cluster, the YEARS= statement specifies the number of years preceding the survey for which the mortality rates should be calculated, and the WEIGHT= statement names the sample weight variable. The BYVAR= statement, which is optional, designates the subclasses for which separate estimates are requested. In addition to the sample cluster, weight, and subclass identification variables, the data set should also contain the following variables:

CASEID	: Respondent identification variable
V008	: Date of interview (CMC)
V011	: Date of birth of the respondent (CMC)
V201	: Total number of children ever born alive
B3	: Child's date of birth (CMC)
B5	: Survival status of the child at the time of interview (1=alive, 0=dead)
B7	: Age at death of the child in completed months, and

Any variable that will be used to identify the subclasses in the BYVAR= statement.

Example 6

***LIBNAME** in 'J:\USER\';*
***OPTIONS MSTORED SASMSTORE**= in;*

%MORT (DATA = child, **PSU**= v021, **YEARS** = 10, **BYVAR** = v102, **WEIGHT** = rweight);

As in previous examples, the LIBNAME and OPTIONS statements bring into the SAS session the MORT macro, which is stored in the directory called 'J:\USER'. The output file is presented below.

Output 6 : Sampling errors of the mortality rates for the last 10 years preceding the survey

SAMPLING ERRORS

V102=.

VARIABLE	MORTRATE	TYPE	RATE	STDERROR	N_UNWGT	N_WGT	SRS	DEFT	RELERROR	LOWER	UPPER
NEOMORT	NEONATAL	Rate	40.506	2.535	11605	12024	1.830	1.385	0.063	35.435	45.577
PNMORT	POSTNEONATAL	Rate	58.792	3.882	11285	10895	2.214	1.753	0.066	51.028	66.556
INMORT	INFANT	Rate	99.298	4.777	11285	10895	2.815	1.697	0.048	89.744	108.853
CMORT	CHILD	Rate	71.670	3.851	9577	8380	2.636	1.461	0.054	63.969	79.372
U5MORT	UNDER 5	Rate	163.852	6.629	9577	8380	3.782	1.753	0.040	150.593	177.111

V102=Urban

VARIABLE	MORTRATE	TYPE	RATE	STDERROR	N_UNWGT	N_WGT	SRS	DEFT	RELERROR	LOWER	UPPER
NEOMORT	NEONATAL	Rate	37.367	5.094	2658	2504	3.679	1.385	0.136	27.179	47.554
PNMORT	POSTNEONATAL	Rate	40.515	5.381	2597	2352	3.869	1.391	0.133	29.752	51.277
INMORT	INFANT	Rate	77.881	6.732	2597	2352	5.259	1.280	0.086	64.416	91.346
CMORT	CHILD	Rate	53.335	5.273	2429	1894	4.559	1.157	0.099	42.789	63.882
U5MORT	UNDER 5	Rate	127.063	8.195	2429	1894	6.758	1.213	0.064	110.673	143.452

V102=Rural

VARIABLE	MORTRATE	TYPE	RATE	STDERROR	N_UNWGT	N_WGT	SRS	DEFT	RELERROR	LOWER	UPPER
NEOMORT	NEONATAL	Rate	41.332	2.915	8947	9519	2.104	1.385	0.071	35.503	47.161
PNMORT	POSTNEONATAL	Rate	63.671	4.646	8633	8543	2.628	1.768	0.073	54.378	72.963
INMORT	INFANT	Rate	105.003	5.716	8633	8543	3.299	1.732	0.054	93.571	116.435
CMORT	CHILD	Rate	76.866	4.607	7148	6486	3.151	1.462	0.060	67.652	86.080
U5MORT	UNDER 5	Rate	173.797	7.927	7148	6486	4.482	1.769	0.046	157.943	189.652

If the child data file is not readily available, one can create it from the woman data file using the CHILD macro, which is attached to the MORT macro. In example 7, the CHILD macro is executed first to create a temporary child data file called *child* using the *test* data file. The created child data file is then used as the input data set for the MORT macro.

Example 7

***LIBNAME** in 'J:\USER\';*
***OPTIONS MSTORED SASMSTORE**= in;*

%CHILD (DATA = test, **OUTDSN** = child);
%MORT (DATA = child, **PSU**= v021, **YEARS** = 5, **BYVAR** = v102, **WEIGHT** = rweight);

5 Interpretation of results

As stated earlier, the estimated sampling error produced by the MEAN macro is biased and the degree of this bias depends on the coefficient of variation of cluster sizes (CV). The estimation of the sampling error using the Taylor series approximation method assumes that a reasonable control of the coefficient of variation of cluster sizes is maintained. Generally, the CV is small in most large surveys; a value of the CV under 0.2 is considered to be acceptable. The printed value of the CV by the MEAN macro allows the validity of the assumption underlying the Taylor variance approximation to be checked.

The relative error of the sample estimate printed in each macro is obtained by dividing the sampling error of the estimate by the sample estimate. In other words, the relative error of an estimate is the proportion of error in the estimated sample value. The relative error is sometimes used to measure the precision of an estimate. However, caution should be exercised since the relative error is large and unstable when the sample estimate is close to zero.

The design effect, as calculated in the MEAN and MORT macros, is the ratio of the estimated sampling error under the actual sample design to the sampling error that would be obtained under simple random sampling. It is considered undefined in each macro if the sampling error under simple random sampling is zero. The design effect is a factor that summarizes the effects of the complexities of the sample design, particularly the effects of clustering and stratification. A value of 1 for the design effect is an indication that the sample design is as efficient as simple random sampling, while a value greater than 1 corresponds to an increase in the sampling error due to the complexity of the sample design.

References

Kish, L. 1965. *Survey Sampling.* New York: Wiley.

Mboup, G., and Saha, T. 1998. *Fertility levels, trends, and differentials.* DHS Comparative Studies No. 28. Calverton, Maryland: Macro International Inc.

Rutsein, S. O. 1984. *Infant and child mortality: Levels, trends, and demographic differentials.* WFS Comparative Studies No. 43. Voorburg, Netherlands: International Statistical Institute.

Thiam, M., and Aliaga, A. 1999. *An SAS macro for estimating the sampling errors of means and proportions in the context of stratified multistage cluster sampling.* Proceedings of the Section on Survey Research Methods, American Statistical Association, 801-806.

Verma, V., and Pearce, M. 1986. *CLUSTERS: A package program for the computation of sampling errors for clustered samples.* Version 3.0. Research Center: International Statistical Institute.

Estimation des erreurs de sondage

des moyennes, de l'indice synthéthique de fécondité et des quotients de mortalité des enfants, en utilisant SAS

MEASURE *DHS+*

Estimation des erreurs de sondage des moyennes, de l'indice synthétique de fécondité et des quotients de mortalité des enfants en utilisant SAS

Mamadou Thiam et Alfredo Aliaga

ORC Macro
Calverton, Maryland

Juin 2001

Remerciements

Les auteurs tiennent à remercier Gora Mboup, Noureddine Abderrahim et Shea Rutstein pour leur importante contribution. Nous tenons également à remercier Monique Barrère d'avoir bien voulu assurer la traduction de ce document en langue française.

Introduction

Les enquêtes réalisées dans les pays en voie de développement dans le cadre du Programme des Enquêtes Démographiques et de Santé (EDS) ont pour principal objectif de recueillir des informations sur la planification familiale et la santé maternelle et infantile. Comme dans beaucoup d'enquêtes, l'échantillon des enquêtes EDS est un échantillon probabiliste basé sur un plan de sondage complexe, caractérisé à la fois par des probabilités inégales de sélection, une stratification, un tirage en grappes et une sélection de différents types d'unités de sondage. Les estimations obtenues à partir d'une enquête par sondage sont sujettes à deux types d'erreurs : les erreurs de réalisation et les erreurs de sondage. Les erreurs de réalisation, difficiles à évaluer, proviennent généralement d'erreurs effectuées lors de la collecte et de l'exploitation des données. Les erreurs de sondage, qui sont celles qui nous intéressent ici, sont strictement dues au fait que l'on étudie qu'une partie de la population. Quand on tire un échantillon probabiliste, celui-ci n'est qu'un échantillon parmi de nombreux autres qui auraient pu être sélectionnés à partir de la même population et en utilisant le même plan de sondage. Chacun de ces échantillons fournirait des estimations qui diffèreraient quelque peu des estimations obtenues à partir de l'échantillon réellement sélectionné. L'erreur de sondage d'une estimation est donc une mesure de la variabilité des estimations obtenues à partir de tous les échantillons possibles, de même taille et conçus dans les mêmes conditions essentielles d'enquête. Le degré de cette variation peut être estimé à partir de l'échantillon probabiliste qui est sélectionné. Les caractéristiques du plan de sondage influent sur l'importance des erreurs de sondage et doivent être prises en compte dans l'estimation des erreurs de sondage.

Ce rapport technique a pour but de présenter trois macros commandes en SAS, d'utilisation simple, pour le calcul des erreurs de sondage : MEAN, TFR et MORT. La macro commande MEAN calcule les erreurs de sondage des moyennes et des proportions alors que les macros commandes TFR et MORT les estiment, respectivement, pour l'indice synthétique de fécondité et les quotients de mortalité des enfants. Ces trois macros commandes produisent les erreurs de sondage non seulement pour tout l'échantillon, mais aussi pour des sous-groupes déterminés. En plus d'estimer les erreurs de sondage, les macros commandes produisent les tailles d'échantillon pondéré et non pondéré, l'erreur de sondage sous un échantillon aléatoire simple, l'effet de grappe (sauf pour l'Indice Synthétique de Fécondité), l'erreur relative et l'intervalle de confiance à 95 % pour l'estimation. Les enquêtes à grande échelle produisent un grand nombre d'estimations mais, dans un rapport d'enquête, les erreurs de sondage peuvent être seulement calculées et présentées pour certaines variables sélectionnées. La macro commande MEAN permettra à l'analyste d'estimer facilement les erreurs de sondage pour les moyennes et les proportions de certaines variables présentant un intérêt dans le cas d'un plan de sondage complexe. Les macros commandes TFR et MORT fournissent aux utilisateurs de SAS un outil supplémentaire qui n'est pas encore disponible dans la version actuelle de SAS. Chaque macro commande est décrite en détail dans les sections suivantes.

2 La macro commande MEAN

2.1 Méthode de calcul et formule

La macro commande MEAN utilise la méthode de linéarisation de Taylor pour estimer les erreurs de sondage des moyennes et proportions à partir d'un modèle d'échantillonnage de grappes finales. Dans le cas du modèle d'échantillon de grappes finales, les éléments à l'intérieur des Unités Primaires de Sondage (UPS) sont d'abord répartis en grappes finales de taille à peu près égale. Ensuite de chaque UPS, un échantillon de grappes finales est sélectionné sans remise. La méthode de linéarisation de Taylor traite chaque estimateur comme un ratio ; elle permet d'obtenir une approximation linéaire de l'estimateur et utilise ensuite la variance de l'approximation pour estimer la variance de l'estimateur lui-même. Cette méthode repose aussi sur l'hypothèse selon laquelle au moins deux grappes sont sélectionnées

indépendemment et avec remise dans chaque strate. En pratique, les grappes sont sélectionnées sans remise et les estimations de la variance obtenues en utilisant la méthode de linéarisation sont alors exagérées, mais cela est négligeable si le taux de sondage est faible. Avec la méthode de linéarisation de Taylor, les quantités agrégées sont seulement calculées au niveau de la grappe; tout ce qui se situe à un niveau plus fin n'est pas utilisé. Il convient de noter que les strates utilisées pour calculer les erreurs de sondage ne sont pas nécessairement les mêmes que celles utilisées dans le plan de sondage et dans la sélection de l'échantillon. Les strates utilisées pour le calcul des erreurs de sondage peuvent être créées en regroupant les grappes sur la base de certaines caractéristiques communes de manière à conserver une certaine homogénéité à l'intérieur de chaque strate. Dans les enquêtes EDS, chaque strate est formée en regroupant 2 ou 3 grappes, et donc en affinant davantage la stratification de départ utilisée dans le plan de sondage.

Pour la formule de base, considérons un ratio r = y/x où y et x sont des sommes pondérées pour l'échantillon total ou un sous-groupe de l'échantillon. Supposons qu'il y ait H strates dans l'échantillon ou le sous-groupe et que m_h grappes sont sélectionnées à partir de la strate h. Pour toute grappe i dans la strate h, on a :

y_{hij} = valeur de la variable y pour l'élément j dans la grappe i de la strate h;
y_{hi} = somme pondérée de toutes les valeurs y_{hij} pour tous les éléments de la grappe i, c'est-à-dire :

$$y_{hi} = \sum_{j=1}^{n_{hi}} w_{hij} y_{hij}$$

où w_{hij} est le coefficient de pondération pour l'élément j dans la grappe i de la strate h, et n_{hi} est le nombre d'éléments dans la grappe i.

y_h = somme des valeurs de y_{hi} pour la strate h, c'est-à-dire :

$$y_h = \sum_{i=1}^{m_h} y_{hi}$$

y = somme pour tout l'échantillon ou le sous-groupe, c'est-à-dire :

$$y = \sum_{h=1}^{H} y_h$$

Des notations identiques s'appliquent à la variable x.

La variance du ratio r est calculée en utilisant la formule ci-dessous, avec l'erreur de sondage qui est la racine carrée de la variance :

$$SE^2(r) = \text{var}(r) = \frac{1}{x^2} \sum_{h=1}^{H} \left[\frac{(1-f_h)m_h}{m_h - 1} \left(\sum_{i=1}^{m_h} z_{hi}^2 - \frac{z_h^2}{m_h} \right) \right] \qquad (1)$$

dans laquelle, $z_{hi} = y_{hi} - r.x_{hi}$ et $z_h = y_h - r.x_h$

où f_h est le taux de sondage de la strate qui est si faible, en particulier pour des échantillons de grande taille qu'on peut ne pas en tenir compte. Notez que x_{hij} est égal à 1 dans la macro commande MEAN.

L'estimation de l'erreur de sondage selon la méthode de linéarisation de Taylor n'est pas exempte de biais. L'importance du biais dépend du coefficient de variation (CV) des tailles des grappes finales et peut être ignoré pour des variations inférieures à 0,2. La macro commande MEAN fournit le coefficient de variation des tailles des grappes pour l'ensemble de l'échantillon et pour chaque sous-groupe de l'échantillon. Ce CV est calculé en utilisant la formule suivante :

$$CV = \frac{1}{x}\left[\sum_{h=1}^{H}\frac{(1-f_h)m_h}{m_h-1}\left(\sum_{i=1}^{m_{hi}}x_{hi}^2 - \frac{(\sum_{i=1}^{m_{hi}}x_{hi})^2}{m_h}\right)\right]^{\frac{1}{2}} \qquad (2)$$

où x_{hi} est la somme des pondérations pour tous les éléments de la grappe i et la strate h.

L'effet de grappe (DEFT) fourni par la macro commande MEAN est défini dans les enquêtes EDS comme le ratio de l'erreur de sondage estimée dans l'échantillon réel par rapport à l'erreur de sondage qui serait obtenue à partir d'un échantillon aléatoire simple. En supposant un sondage aléatoire simple, l'erreur de sondage pour le ratio r = y/x est donnée par la formule suivante :

$$SE_{srs}^2(r) = \frac{1}{n-1}\frac{\sum_{h=1}^{H}(1-f_h)\sum_{i=1}^{m_h}\sum_{j=1}^{n_{hi}}w_{hij}z_{hij}^2}{\sum_{h=1}^{H}\sum_{i=1}^{m_h}\sum_{j=1}^{n_{hi}}w_{hij}} \qquad (3)$$

où $z_{hij} = y_{hij} - r.x_{hij}$

2.2 Spécification des paramètres

Le groupe d'instructions suivant, dans lequel les paramètres clé restent à définir, permet d'exécuter la macro commande MEAN en SAS :

%MEAN (DATA=, WEIGHT=, PSU=, STRATA=, BYVAR=, ALLVARS=);

Dans le groupe d'instructions ci-dessus, l'instruction *DATA=* désigne le nom du fichier de données qui contient toutes les variables pour lesquelles l'erreur de sondage doit être calculée ainsi que toutes les autres variables identifiées dans les instructions *WEIGHT, PSU, STRATA, et BYVAR.* L'instruction *WEIGHT=* désigne la variable identifiant le coefficient de pondération. L'instruction *PSU=* identifie l'unité primaire de sondage à laquelle chaque grappe finale de l'échantillon appartient, alors que l'instruction *STRATA=* identifie la variable de stratification qui doit être utilisée pour l'estimation des erreurs de sondage. Si les erreurs de sondage doivent être estimées pour des sous-groupes de

l'échantillon, la variable les définissant doit être identifiée dans l'instruction *BYVAR=*. Toutes les variables pour lesquelles les erreurs de sondage seront calculées doivent être énumérées et séparées par un espace dans l'instruction *ALLVARS=*. Il faut noter qu'à l'exception de l'instruction *BYVAR=* qui est optionelle, toutes les autres instructions sont nécessaires pour une exécution correcte de la macro commande. Si des estimations sont demandées pour des sous-groupes de l'échantillon, la macro commande produit les estimations pour ces sous-groupes de l'échantillon comme pour l'échantillon total.

2.3 Fichier de sortie

La macro commande MEAN crée et imprime un fichier de données en SAS contenant les résultats suivants pour chaque variable pour laquelle l'erreur de sondage a été calculée :

CV : Coefficient de variation des tailles de grappes

VARIABLE : Nom de la variable

LABEL : Libellé de la variable, s'il en existe (40 caractères au maximum, y compris les espaces)

TYPE : Type de statistique (moyenne ou proportion)

MEAN : Moyenne ou proportion dans l'échantillon pondéré

STDERROR : Erreur de sondage sur l'estimation de la moyenne ou de la proportion

N_UNWGT : Nombre de cas non pondéré utilisé dans les calculs

N_WGT : Nombre de cas pondéré utilisé dans les calculs

SRS : Erreur de sondage de la moyenne ou proportion sous un échantillon aléatoire simple

DEFT : Effet de grappe (rapport des erreurs de sondage)

RELERROR : Erreur relative de la moyenne ou de la proportion dans l'échantillon

LOWER : Limite inférieure de l'intervalle de confiance à 95%

UPPER : Limite supérieure de l'intervalle de confiance à 95%

Les estimations obtenues en utilisant la macro commande MEAN ont été comparées avec celles obtenues par la méthode de linéarisation de Taylor du logiciel SUDAAN et du module d'erreur de sondage du logiciel « Integrated System for Survey Analysis » (ISSA). Les trois programmes ont donné les mêmes estimations.

Exemples

Les exemples présentés dans ce rapport utilisent les données de l'Enquête Démographique et de Santé réalisée à Madagascar en 1997. L'enquête avait, entre autres, pour objectif l'estimation du niveau de la prévalence contraceptive chez les femmes de 15-49 ans, l'estimation du niveau de la fécondité et des niveaux de mortalité des enfants. À titre d'illustration, figurent ci-dessous les variables utilisées :

V102 : type de résidence (1 = urbain, 2 = rural)

V106 : plus haut niveau d'instruction atteint (0 = aucune instruction, 1 = primaire,
2 = secondaire, 3 = supérieure)
V201 : nombre total d'enfants nés vivants des femmes
V502 : état matrimonial (0= jamais marié, 1 = actuellement marié, 2 = précédemment marié)
V613 : nombre idéal d'enfants
V005 : coefficient de pondération

Un échantillon de 7 060 femmes avec interview complète a été obtenu en utilisant un sondage stratifié à deux degrés. Au premier degré, 270 zones de dénombrement (ZD) ont été sélectionnées. Dans ces ZD, on a procédé à une énumération complète des ménages. La liste des ménages obtenue a servi de base de sondage pour la sélection des ménages au second degré. Toutes les femmes éligibles identifiées dans les ménages sélectionnés ont été incluses dans l'échantillon.

Les variables nominales doivent être recodées pour estimer les proportions qui ont un intéret pour l'enquête. Les instructions suivantes créent le fichier de travail TEST qui contiendra les nouvelles variables devant être utilisées dans la macro commande. Il faut noter que dans le fichier de données d'origine (MDIR), la valeur des coefficients de pondération de l'échantillon est multipliée par 1 000 000.

```
/***** RECODING VARIABLES ******/

DATA TEST; SET MDIR;
 RWEIGHT=V005/1000000;          /* variable pour le coefficient de pondération*/
 EVBORN = V201;
 IF V102=1 THEN URBAN=1; ELSE URBAN=0;
 IF V106 IN (2,3) THEN EDUC=1; ELSE EDUC=0;
 IF 0<=V613<45 THEN IDEAL=V613; ELSE IDEAL=.;
 IF V502=1 THEN CURMAR=1; ELSE CURMAR=0;
 IF 1<=V313 <=3 THEN CUSE=1; ELSE CUSE=0 ;
 IF CUMAR=0  THEN CUSE=. ;

 LABEL URBAN     = 'Résidence urbaine'
       EDUC      = 'Avec instruction secondaire ou plus'
       CURMAR    = 'Actuellement marié (en union)'
       CUSE      = 'Utilise une méthode'
       EVBORN    = 'Enfants déjà nés'
       IDEAL     = 'Nombre idéal d'enfants';

RUN;
```

Exemple 1 : Calcul des erreurs de sondage pour des sous groupes de l'échantillon n'est pas demandé.

```
LIBNAME in 'J:\USER\';
OPTIONS MSTORED SASMSTORE= in;

%MEAN (DATA= test, WEIGHT= rweight, PSU=v021, STRATA=v022,
       ALLVARS= urban educ cuse evborn ideal);
```

Dans cet exemple, les instructions LIBNAME et OPTIONS font appel à la macro commande MEAN, qui est mémorisée dans le répertoire 'J:\USER'. Les erreurs de sondage seront produites quand l'instruction %MEAN sera exécutée.

Exemple 2 : Les erreurs de sondage sont demandées pour des sous groupes de l'échantillon.

```
LIBNAME in  'J:\USER\';
OPTIONS MSTORED SASMSTORE= in;

%MEAN (DATA=test, WEIGHT= rweight, PSU=v021, STRATA=v022,
        BYVAR=v102, ALLVARS=urban educ cuse evborn ideal);
```

Les résultats obtenus pour ces deux exemples sont présentés ci-dessous.

Résultats de l'exemple 1 : Les erreurs de sondage pour des sous groupes de l'échantillon ne sont pas demandées

COEFFICIENT DE VARIATION DES TAILLES DE GRAPPE: ÉCHANTILLON TOTAL

VARIABLE	CV
CUSE	0.030
EDUC	0.029
EVBORN	0.029
IDEAL	0.029
URBAN	0.029

ERREURS DE SONDAGE: ÉCHANTILLON TOTAL

VARIABLE	LABEL	TYPE	MEAN	STDERROR	N_UNWGT	N_WGT	SRS	DEFT	RELERROR	LOWER	UPPER
CUSE	Utilise une méthode	Proportion	0.194	0.010	4356	4435	0.006	1.668	0.051	0.174	0.214
EDUC	Instruction secondaire ou plus	Proportion	0.269	0.012	7060	7060	0.005	2.310	0.045	0.244	0.293
EVBORN	Enfants déjà nés	Moyenne	3.215	0.050	7060	7060	7060	0.038	1.308	0.016	3.115
IDEAL	Nombre idéal d'enfants	Moyenne	5.306	0.082	6447	6358	0.035	2.303	0.015	5.142	5.469
URBAN	Résidence urbain	Proportion	0.281	0.011	7060	7060	0.005	2.007	0.038	0.260	0.303

Résultats de l'exemple 2 : Les erreurs de sondage sont demandées pour 2 sous groupes de l'échantillon (urbain et rural)

COEFFICIENT DE VARIATION DES TAILLES DE GRAPPE: ÉCHANTILLON TOTAL

VARIABLE	CV
CUSE	0.030
EDUC	0.029
EVBORN	0.029
IDEAL	0.029
URBAN	0.029

ERREURS DE SONDAGE :ÉCHANTILLON TOTAL

VARIABLE	LABEL	TYPE	MEAN	STDERROR	N_UNWGT	N_WGT	SRS	DEFT	RELERROR	LOWER	UPPER
CUSE	Utilise une méthode	Proportion	0.194	0.010	4356	4435	0.006	1.668	0.051	0.174	0.214
EDUC	Instruction secondaire ou plus	Proportion	0.269	0.012	7060	7060	0.005	2.310	0.045	0.244	0.293
EVBORN	Enfants déjà nés	Moyenne	3.215	0.050	7060	7060	0.038	1.308	0.016	3.115	3.315
IDEAL	Nombre idéal d'enfants	Moyenne	5.306	0.082	6447	6358	0.035	2.303	0.015	5.142	5.469
URBAN	Résidence urbain	Proportion	0.281	0.011	7060	7060	0.005	2.007	0.038	0.260	0.303

COEFFICIENT DE VARIATION DES TAILLES DE GRAPPE : SOUS GROUPES DE L'ÉCHANTILLON

---------- V102=Urbain ----------

VARIABLE	CV
CUSE	0.050
EDUC	0.038
EVBORN	0.038
IDEAL	0.039
URBAN	0.038

---------- V102=Rural ----------

VARIABLE	CV
CUSE	0.036
EDUC	0.037
EVBORN	0.037
IDEAL	0.037
URBAN	0.037

ERREURS DE SONDAGE : SOUS GROUPES DE L'ÉCHANTILLON

---------- V102=Urbain ----------

VARIABLE	LABEL	TYPE	MEAN	STDERROR	N_UNWGT	N_WGT	SRS	DEFT	RELERROR	LOWER	UPPER
CUSE	Utilise une méthode	Proportion	0.345	0.022	1307	1132	0.013	1.680	0.064	0.301	0.390
EDUC	Instruction secondaire ou plus	Proportion	0.526	0.031	2376	1987	0.010	3.043	0.059	0.463	0.588
EVBORN	Enfants déjà nés	Moyenne	2.496	0.070	2376	1987	0.057	1.225	0.028	2.356	2.636
IDEAL	Nombre idéal d'enfants	Moyenne	4.181	0.115	2255	1857	0.046	2.469	0.027	3.951	4.410
URBAN	Résidence urbain	Proportion	1000	0.000	2376	1987	0.000	.	0.000	1.000	1.000

---------- V102=Rural ----------

VARIABLE	LABEL	TYPE	MEAN	STDERROR	N_UNWGT	N_WGT	SRS	DEFT	RELERROR	LOWER	UPPER
CUSE	Utilise une méthode	Proportion	0.143	0.011	3049	3302	0.006	1.789	0.079	0.120	0.165
EDUC	Instruction secondaire ou plus	Proportion	0.168	0.010	4684	5073	0.005	1.884	0.061	0.147	0.189
EVBORN	Enfants déjà nés	Moyenne	3.496	0.061	4684	5073	0.048	1.260	0.017	3.374	3.619
IDEAL	Nombre idéal d'enfants	Moyenne	5.770	0.105	4192	4501	0.046	2.306	0.018	5.560	5.980
URBAN	Résidence urbain	Proportion	0.000	0.000	5073	5073	0.000	.	.	0.000	0.000

2.4 Extension de la macro commande MEAN

La macro commande MEAN peut également être utilisée pour calculer les erreurs de sondage de proportions plus complexes. Supposons que l'on souhaite estimer la proportion d'enfants de moins de 3 ans qui ont eu la diarrhée durant les deux dernières semaines et qui ont reçu un traitement. Dans cet exemple, pour pouvoir utiliser la macro commande MEAN, il est nécessaire de recoder les variables en 2 étapes, comme présentés ci-dessous dans le programme SAS.

Exemple 3

```
Data test2; set child;
                                        /**** Première étape ****/

if b5 = 0 or age_chld >= 36 then diar2w = .;
if b5 = 1 and h11 = 2 and age_chld <= 35 then diar2w = 1;
if b5 = 1 and h11^= 2 and age_chld <= 35 then diar2w = 0;

                                        /**** Deuxième étape ***/

if diar2w = 1 and h12z = 1 then medtre = 1;
if diar2w = 1 and h12z ^= 1 then medtre = 0;

label diar2w    = 'A eu la diarrhée durant les deux dernières semaines'
  age_chld      = 'Âge de l'enfant'
      b5        = 'Enfant vivant'
     h11        = 'Enfant a eu la diarrhée dans les dernières 24 heures ou pendant les 2
                   dernières semaines, mais pas dans les dernières 24 heures'
     h12z       = 'Enfant conduit dans un établissement médical pour traitement contre la
                   diarrhée'
   medtre       = 'A recherché un traitement médical';
Run;
```

Dans la première étape, la variable *diar2w* est créée pour identifier les enfants vivants de moins de 36 mois ayant eu la diarrhée durant les 2 semaines ayant précédé l'enquête. Parmi ces enfants, certains ont reçu un traitement médical et sont identifiés ultérieurement dans la seconde étape en utilisant la variable *medtre*. Une fois données, les instructions suivantes permettent le calcul de l'erreur de sondage.

```
LIBNAME in 'J:\USER\';
OPTIONS MSTORED SASMSTORE= in;

%MEAN (DATA=test2, WEIGHT= rweight, PSU=v021, STRATA=v022,
            BYVAR=v102, ALLVARS=medtre);
```

3 La macro commande TFR

3.1 Méthode de calcul et formule

Le calcul de l'Indice Synthétique de Fécondité (ISF) est basé sur l'historique complet des naissances de la femme collecté dans la section reproduction du questionnaire EDS. L' ISF, pour une certaine période précédant l'enquête, est calculé dans les enquêtes EDS en utilisant des groupes d'âges

quinquennaux : 15-19, 20-24 , 25-29, 30-34, 35-39, 40-44 et 45-49 ans. Pour la même période, le nombre de naissances vivantes des femmes dans chaque groupe d'âges et le nombre de femmes-années d'exposition au risque de grossesse dans le même groupe d'âges sont utilisés pour calculer l'Indice Synthétique de Fécondité comme suit.

$$ISF(t) = 5 \times \sum_i \frac{B(i,t)}{E(i,t)}$$

où i est le groupe d'âges, B(i,t) le nombre de naissances vivantes de la femme dans le groupe d'âges i durant la période t, et E(i,t) le nombre de femmes-années d'exposition au risque de grossesse parmi les femmes du groupe d'âges i durant la période t. Il faut noter que les naissances survenues durant le mois de l'enquête sont exclues.

En calculant B(i,t), la macro commande TFR crée à partir du fichier de données femme, un fichier temporaire de données enfant et une variable qui permet de compter les naissances. Pour un groupe d'âges particulier i = [a, b], cette variable est positionnée à 1 quand la naissance a lieu durant la période en question et quand l'âge de la mère à la naissance se situe entre a et b; sinon, elle est positionnée à 0. Par conséquent, B(i,t) est obtenu en prenant la somme pondérée des valeurs pour cette variable comme suit :

$$B(i,t) = \sum_j w_j I_{[a,b]}$$

où $I_{[a,b]} = 1$ si $a \leq$ MOM_AGE $\leq b$ et $0 <$ CHD_AGE $\leq t$
$= 0$ autrement

et MOM_AGE est l'âge de la mère à la naissance, CHD_AGE est l'âge de l'enfant à l'enquête, et w_j est le coefficient de pondération de l'échantillon pour les femmes.

Le nombre de femmes-années d'exposition au risque de grossesse parmi les femmes du groupe d'âges i = [a, b] durant la période t dépend de l'âge de la mère à l'enquête (AGE_MAX), et de l'âge de la mère au début de la période en question (AGE_MIN). E(i,t) est calculé comme suit :

$$E(i,t) = \sum_j w_j EXP\_ab$$

où w_j est le coefficient de pondération de l'échantillon pour les femmes, et

EXP_ab = [b – maximum(a, AGE_MIN)]/12 si AGE_MIN < b < AGE_MAX

= [AGE_MAX – maximum(a, AGE_MIN)]/12 si a < AGE_MAX < b

Dans les enquêtes qui comprennent seulement les femmes non-célibataires, le dénominateur dans le calcul de l'ISF est augmenté pour inclure toutes les femmes. Le facteur d'augmentation est la proportion des femmes non-célibataires de 15-49 ans calculée à partir du tableau ménage.

La macro commande TFR utilise la méthode de Jackknife pour estimer l'erreur de sondage de l'Indice Synthétique de Fécondité (ISF). L'ISF est estimé à partir de chacun des sous-échantillons de l'échantillon principal. Chaque sous-échantillon exclut une grappe dans le calcul de l'estimation. La

variabilité des estimations faites à partir des sous-échantillons est alors utilisée pour calculer l'erreur de sondage de l'ISF comme le montre la formule ci-dessous :

$$SE^2(ISF) = \text{var}(ISF) = \frac{1}{k(k-1)} \sum_{i=1}^{k} (r_i - r)^2 \qquad (4)$$

où : $r_i = k\,r - (k-1)\,r_{(i)}$

et
- k est le nombre total de grappes,
- r est l'estimation calculée à partir de l'échantillon total de k grappes, et
- $r_{(i)}$ est l'estimation calculée à partir du sous-échantillon qui exclut la $i^{ème}$ grappe.

3.2 Spécification des paramètres

Le groupe d'instructions suivant, dans lequel les valeurs des paramètres clé restent à définir, permet d'exécuter la macro commande TFR en SAS :

%TFR (DATA =, PSU=, YEARS =, BYVAR=, WEIGHT =, INFLATE=);

où DATA est le nom du fichier de données SAS, PSU identifie la grappe, YEARS est le nombre d'années précédant l'enquête, BYVAR est la variable identifiant les sous groupes de l'échantillon pour lesquels des résultats sont demandés séparément et WEIGHT est la variable identifiant le coefficient de pondération. L'instruction INFLATE= désigne le facteur d'inflation qui doit être utilisé pour les échantillons de femmes non-célibataires. À l'exception des instructions INFLATE= et BYVAR=, toutes les autres instructions sont obligatoires. Si les deux instructions BYVAR= et INFLATE= sont spécifiées à la fois, la macro commande TFR produira des résultats uniquement pour chaque niveau de la variable spécifiée dans l'instruction BYVAR=. Si l'instruction BYVAR= est spécifiée et INFLATE= est omise, la macro commande produira des résultats pour l'échantillon total et pour chaque niveau de la variable spécifiée dans l'instruction BYVAR=.

En plus des variables utilisées dans les instructions PSU=, WEIGHT=, BYVAR= et INFLATE=, le fichier de données SAS doit également contenir les variables suivantes :

CASEID : variable d'identification de l'enquêté
V008 : date de l'interview (CMC)
V011 : date de naissance de l'enquêté (CMC)
V201 : nombre total d'enfants nés vivants
B3 : date de naissance de l'enfant (CMC)
B5 : état de survie de l'enfant au moment de l'enquête (recodé tel que 1 = vivant, 0 = décédé)
B7 : âge au décès de l'enfant (CMC), et
Toute variable qui sera utilisée pour identifier les sous groupes de l'échantillon dans l'instruction BYVAR=.

La macro commande TFR crée et imprime un fichier de données contenant l'Indice Synthétique de Fécondité et son erreur de sondage, le nombre de cas pondérés, l'erreur relative et les limites inférieure et supérieure de l'intervalle de confiance à 95%. Pour l'indice Synthétique de Fécondité, le nombre de cas non pondéré n'est pas pertinent étant donné qu'il n'y a aucune valeur non pondérée connue pour les femmes-années d'exposition au risque de grossesse.

Exemple 4

Comme dans l'exemple 1, le même fichier de données femme (*test*) est utilisé, et les erreurs de sondage pour les sous groupes de l'échantillon ne sont pas demandées.

***LIBNAME** in 'J:\USER\';*
***OPTIONS MSTORED SASMSTORE**= in;*

***%TFR (DATA** = test, **PSU** = v021, **YEARS** = 3, **WEIGHT** = rweight);*

Exemple 5

Dans cet exemple, les erreurs de sondage sont demandées pour des sous groupes de l'échantillon.

***LIBNAME** in 'J:\USER\';*
***OPTIONS MSTORED SASMSTORE**= in;*

***%TFR (DATA = test, PSU = v021, YEARS = 3, BYVAR = v102, WEIGHT** = rweight);*

Les résultats pour ces 2 exemples sont présentés ci-dessous.

Résultats de l'exemple 4: Erreurs de sondage pour des sous groupes de l'échantillon ne sont pas demandés

---------- TOTAL=échantillon total----------

VARIABLE	LABEL	TYPE	TFR	STDEROR	N_UWGT	N_WGT	RELERROR	LOWER	UPPER
TFR	ISF	Rate	5.969	0.140	NA	19801	0.023	5.689	6.249

Résultats de l'exemple 5: Erreurs de sondage pour 2 sous groupes de l'échantillon sont demandés (urbain et rural)

---------- TOTAL=Échantillon total ----------

VARIABLE	LABEL	TYPE	TFR	STDEROR	N_UWGT	N_WGT	RELERROR	LOWER	UPPER
TFR	ISF	Rate	5.969	0.140	NA	19801	0.023	5.689	6.249

---------- V102=Urbain ----------

VARIABLE	LABEL	TYPE	TFR	STDEROR	N_UWGT	N_WGT	RELERROR	LOWER	UPPER
TFR	ISF	Rate	4.190	0.029	NA	5554	0.050	3.771	4.609

---------- V102=Rural ----------

VARIABLE	LABEL	TYPE	TFR	STDEROR	N_UWGT	N_WGT	RELERROR	LOWER	UPPER
TFR	ISF	Rate	6.662	0.153	NA	14246	0.023	6.356	6.967

4 La macro commande MORT

4.1 Méthode de calcul et formule

Pour les périodes de 5 ou 10 années qui précédent l'enquête, la macro commande **MORT** calcule l'erreur de sondage pour les quotients de mortalité des enfants suivants :

Quotient de mortalité néonatale : probabilité de décéder entre la naissance et l'âge exact d'un mois;

Quotient de mortalité post-néonatale : probabilité de décéder entre les âges exacts de 1 et 11 mois;

Quotient de mortalité infantile : probabilité de décéder entre la naissance et l'âge exact d'un an;

Quotient de mortalité juvénile : probabilité de décéder entre les âges exacts de 1 et 5 ans;

Quotient de mortalité infanto-juvénile : probabilité de décéder entre la naissance et l'âge exact de 5 ans.

Le calcul de ces quotients dans la macro commande MORT suit la procédure développée par Rutstein (1984). Les probabilités de décéder, durant une période spécifique donnée, sont obtenues à partir des probabilités calculées pour 8 intervalles d'âges : moins d'un mois, 1-2 mois, 3-5 mois, 6-11 mois, 12-23 mois, 24-35 mois, 36-47 mois, et 48-59 mois. La probabilité de décéder pour un intervalle d'âges est définie comme le rapport du nombre de décès qui surviennent chez des enfants exposés au décès dans l'intervalle d'âges, au nombre d'enfants exposés. La probabilité de décéder pour un sous-intervalle est obtenue en soustrayant la probabilité de survie de 1. La macro commande MORT calcule la probabilité de survie pour chaque sous-intervalle et ces probabilités sont aggrégées pour obtenir le quotient de mortalité en utilisant la formule suivante :

$$(n)q(x) = 1 - \prod_{i=x}^{i=x+n}(1 - q(i)) \qquad (5)$$

où (n)q(x) est la probabilité de décéder entre les âges x et x + n, et q(i) est la probabilité de décéder dans le sous-intervalle.

La macro commande MORT fournit une approximation très proche du quotient de mortalité post-néonatale en soustrayant le quotient de mortalité néonatale du quotient de mortalité infantile.

L'erreur de sondage de chaque quotient de mortalité est calculée pour la période des 5 ou 10 ans précédant l'enquête en utilisant la méthode de Jackknife (formule (4)). Pour chaque quotient, le nombre de cas utilisés dans le calcul correspond au nombre d'enfants exposés au risque de décéder entre les âges limites durant la période en question. Pour les quotients néonatal et post-néonatal, le nombre de cas est la valeur minimale des nombres d'enfants exposés au décès pour les intervalles d'âges : moins d'un mois, 1-2 mois, 3-5 mois et 6-11 mois. Pour le quotient de mortalité juvénile, le nombre de cas est la valeur minimale des nombres d'enfants exposés au risque de décéder pour les intervalles d'âges : 12-23 mois, 24-35 mois, 36-47 mois et 48-59 mois. Pour le quotient de mortalité infanto-juvénile, le nombre de cas est la valeur minimale des nombres d'enfants exposés au risque de décéder pour tous les 8 intervalles d'âges. Le nombre de cas non pondéré est utilisé pour calculer l'erreur de sondage sous un échantillon aléatoire simple pour chaque quotient de mortalité en utilisant la formule suivante :

$$SE(quotient) = \left[\frac{quotient(1 - quotient)}{n}\right]^{\frac{1}{2}} \qquad (6)$$

où n est le nombre non pondéré de cas.

4.2 Spécification des paramètres

Pour calculer les erreurs de sondage des quotients de mortalité des enfants, utilisez Le groupe d'instructions suivant en SAS après avoir spécifié la valeur de chaque paramètre :

%MORT (DATA =, PSU =, YEARS =, BYVAR =, WEIGHT =);

L'instruction DATA= identifie le nom du fichier de données enfant, l'instruction PSU= identifie la grappe, l'instruction YEARS= le nombre d'années précédant l'enquête pour lesquelles les quotients de mortalité doivent être calculés, et l'instruction WEIGHT= donne le nom de la variable pour le coefficient de pondération. L'instruction BYVAR= qui est optionnelle désigne les sous groupes de l'échantillon pour lesquels des estimations séparées sont demandées. En plus des variables d'identification de la grappe, du coefficient de pondération et des sous groupes de l'échantillon, le fichier des données doit également contenir les variables suivantes :

CASEID	: variable d'identification de l'enquêté
V008	: date de l'enquête (CMC)
V011	: date de naissance de l'enquêté (CMC)
V201	: nombre total d'enfants nés vivants
B3	: date de naissance de l'enfant (CMC)
B5	: état de survie de l'enfant au moment de l'enquête (recodé tel que 1=vivant, 0 = décédé)
B7	: âge de l'enfant au décès en mois révolus, et

Toute variable qui sera utilisée pour identifier les sous groupes de l'échantillon dans l'instruction BYVAR=.

Exemple 6

***LIBNAME** in 'J:\USER\';*
***OPTIONS MSTORED SASMSTORE**= in;*

***%MORT (DATA** = child, **PSU**= v021, **YEARS** = 10, **BYVAR** = v102, **WEIGHT** = rweight);*

De même que dans les exemples précédents, les instructions LIBNAME et OPTIONS font appel à la macro commande MORT, qui est mémorisée dans le répertoire 'J:\USER'. Les résultats sont présentés ci-dessous.

Résultats de l'exemple 6 : Erreurs de sondage des quotients de mortalité pour la période de 10 ans précédent l'enquête

ERREURS DE SONDAGE

---------- V102=. ----------

VARIABLE	MORTRATE	TYPE	RATE	STDERROR	N_UNWGT	N_WGT	SRS	DEFT	RELERROR	LOWER	UPPER
NEOMORT	NEONATAL	Rate	40.506	2.535	11605	12024	1.830	1.385	0.063	35.435	45.577
PNMORT	POSTNEO	Rate	58.792	3.882	11285	10895	2.214	1.753	0.066	51.028	66.556
INMORT	INFANT	Rate	99.298	4.777	11285	10895	2.815	1.697	0.048	89.744	108.853
CMORT	CHILD	Rate	71.670	3.851	9577	8380	2.636	1.461	0.054	63.969	79.372
U5MORT	UNDER 5	Rate	163.852	6.629	9577	8380	3.782	1.753	0.040	150.593	177.111

---------- V102=Urbain ----------

VARIABLE	MORTRATE	TYPE	RATE	STDERROR	N_UNWGT	N_WGT	SRS	DEFT	RELERROR	LOWER	UPPER
NEOMORT	NEONATAL	Rate	37.367	5.094	2658	2504	3.679	1.385	0.136	27.179	47.554
PNMORT	POSTNEO	Rate	40.515	5.381	2597	2352	3.869	1.391	0.133	29.752	51.277
INMORT	INFANT	Rate	77.881	6.732	2597	2352	5.259	1.280	0.086	64.416	91.346
CMORT	CHILD	Rate	53.335	5.273	2429	1894	4.559	1.157	0.099	42.789	63.882
U5MORT	UNDER 5	Rate	127.063	8.195	2429	1894	6.758	1.213	0.064	110.673	43.452

---------- V102=Rural ----------

VARIABLE	MORTRATE	TYPE	RATE	STDERROR	N_UNWGT	N_WGT	SRS	DEFT	RELERROR	LOWER	UPPER
NEOMORT	NEONATAL	Rate	41.332	2.915	8947	9519	2.104	1.385	0.071	35.503	47.161
PNMORT	POSTNEO	Rate	63.671	4.646	8633	8543	2.628	1.768	0.073	54.378	72.963
INMORT	INFANT	Rate	105.003	5.716	8633	8543	3.299	1.732	0.054	93.571	116.435
CMORT	CHILD	Rate	76.866	4.607	7148	6486	3.151	1.462	0.060	67.652	86.080
U5MORT	UNDER 5	Rate	173.797	7.927	7148	6486	4.482	1.769	0.046	157.943	189.652

Si le fichier de données enfant n'est pas disponible, on peut le créer à partir du fichier de donnés femme en utilisant la macro commande CHILD qui est fournie avec la macro commande MORT. Dans l'exemple 7, la macro commande CHILD est exécutée d'abord pour créer un fichier de données enfant temporaire appelé *child* en utilisant le fichier de données *test*. Le fichier de données enfant créé est alors utilisé comme un fichier de données d'entrée pour la macro commande MORT.

Exemple 7

```
LIBNAME in 'J:\USER\';
OPTIONS MSTORED SASMSTORE= in;

%CHILD (DATA = test, OUTDSN = child);
%MORT (DATA = child, PSU= v021, YEARS = 5, BYVAR = v102, WEIGHT = rweight);
```

5 Interprétation des résultats

Comme nous l'avons mentionné précédemment, l'estimation de l'erreur de sondage produite par la macro commande MEAN est biaisée et l'importance de ce biais dépend du coefficient de variation des tailles de grappes (CV). L'estimation de l'erreur de sondage par la méthode de linéarisation de Taylor suppose qu'un contrôle raisonnable du coefficent de variation des tailles de grappes est maintenu. Généralement, le CV est faible dans la plupart des grandes enquêtes; une valeur du CV inférieure à 0,2 est considérée comme acceptable. La valeur du CV produite par la macro commande MEAN permet de vérifier la validité de l'hypothèse sur laquelle repose l'estimation de la variance par la méthode de Taylor.

L'erreur relative sur l'estimation produite dans chaque macro commande est obtenue en divisant l'erreur de sondage de l'estimation par la valeur de l'estimateur. En d'autres termes, l'erreur relative d'une estimation est la proportion de l'erreur dans la valeur estimée. L'erreur relative est parfois utilisée pour mesurer la précision d'une estimation. Cependant, il est nécessaire d'être prudent puisque l'erreur relative est importante et instable quand la valeur de l'estimateur est proche de zéro.

L'effet de grappe, tel qu'il est calculé dans les macros commande MEAN et MORT, est le rapport de l'erreur de sondage estimée sous le plan de sondage adopté et l'erreur de sondage qui serait obtenue avec un échantillon aléatoire simple. Il est considéré comme indéterminé dans chaque macro commande si l'erreur de sondage sous un échantillon aléatoire simple est égale à zéro. L'effet de grappe est un facteur qui résume les effets de la complexité du plan de sondage, en particulier les effets d'un sondage à plusieurs degrés et de stratification. Une valeur égale à 1 pour l'effet de grappe indique que le plan de sondage est aussi efficace qu'un échantillon aléatoire simple alors qu'une valeur supérieure à 1 indique un accroissement de l'erreur de sondage dû à la complexité du plan de sondage.

Références

Kish, L. 1965. *Survey Sampling.* New York: Wiley.

Mboup, G., et Saha, T. 1998. *Fertility levels, trends and differentials.* DHS Comparative Studies No. 28. Calverton, Maryland: Macro International Inc.

Rutsein, S. O. 1984. *Infant and child mortality: Levels, trends, and demographic differentials.* WFS Comparative Studies No. 43. Voorburg, Netherlands: International Statistical Institute.

Thiam, M., et Aliaga, A. 1999. *An SAS macro for estimating the sampling errors of means and proportions in the context of stratified multistage cluster sampling.* Proceedings of the Section on Survey Research Methods, American Statistical Association, 801-806.

Verma, V., et Pearce, M. 1986. CLUSTERS: *A package program for the computation of sampling errors for clustered samples.* Version 3.0. Research Center: International Statistical Institute.