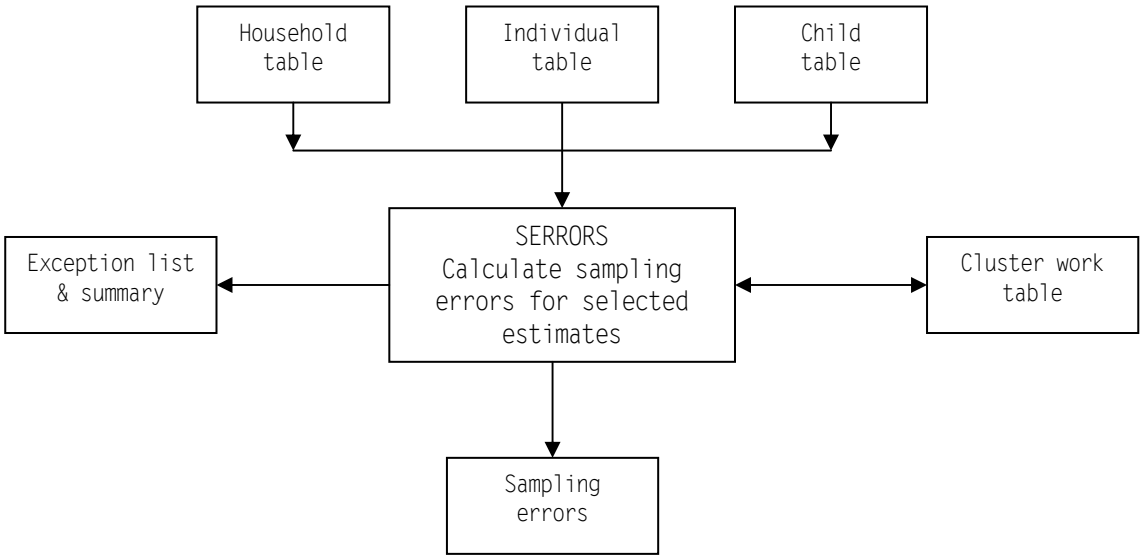
Program: SErrors - Calculate sampling errors using the jack knife method

1. Schematic



2. Function

Calculate sampling errors for the principal estimates presented in the standard CWIQ tables at the national level and for rural, urban and regional subgroups. The list of estimates for which sampling errors are calculated appears in Appendix A.

3. Input

A. CWIQ questionnaire database.

The CWIQ questionnaire database consists of three tables: the household table; the individual table; and the child table. For each household surveyed there is one record in the household table (HHData), one record for each household member in the individual table (INData) and one record for each child under 5 in the child table (CHData).

4. Input/Output

A. Cluster work table (SErrorsW)

The cluster work table contains intermediate results needed to calculate the sampling errors. There is an entry in the cluster work table for each cluster and each estimate for which sampling errors are calculated. A detailed description of the work table appears in Appendix B.

5. Output

A. Exception list and summary (EditList)

This table contains a summary list of the number of questionnaires processed and a list of any exceptional conditions encountered in the data.

B. Sampling errors for selected estimates (SErrors)

This table contains the results of the sampling error calculations. These include the estimate name, the domain (subgroup) name, the estimate itself, the variance, the standard error, the relative standard error and the limits of a 95% confidence interval for the estimate and similar statistics calculated as if the sample was drawn as a simple random sample (SRS). There is a table entry for each national and subgroup estimate. A complete description of the table appears in Appendix B.

6. Processing

This program uses the jack knife method to calculate sampling (standard) errors for ratio estimates of the form r = y / x. For a sample with n clusters, estimates are constructed for n sub-samples, each consisting of n-1 clusters. The n sub-sample estimates are used to calculate the standard error according to the formula below.

The method used to construct the n sub-samples of n-1 clusters is as follows:

For each estimate and each cluster (note that all estimates are weighted):

Accumulate cluster totals for the numerator and denominator of each estimate (y_c and x_c). Write the cluster totals for each estimate to the cluster work table together with the urban/rural and regional classifications for the cluster.

Count the number of clusters (n).

Accumulate domain (national, region, urban/rural) totals for the numerator and denominator of each estimate (y_t and x_t). Calculate domain estimates (r_t) using the accumulated domain totals (y_t, x_t).

Read the cluster work table and for each cluster, calculate totals for an n-1 cluster sample by subtracting the cluster totals of the numerator and denominator from the domain totals (y_i and x_i). Calculate the sample value of the ratio (r_i) using the sample numerator and denominator. Accumulate the sum of the sample estimates.

$y_i = y_t - y_c$; $x_i = x_t - x_c$; $r_i = y_i / x_i$. (i = 1 to n is the sample number)

Calculate and accumulate the sample contribution to the variance of the estimate:

$var_i = (n * r_t - (n-1) * r_i)^2$

The variance of the estimate is calculated using the following formula:

$$\text{Variance} = \frac{\sum_{i=1}^{n} \text{var}_i - n\bar{r}^2}{n(n-1)} \quad \text{where } \bar{r} = \frac{\sum_{i=1}^{n} r_i}{n}$$

The standard error of the estimate is the square root of the variance.

The relative standard error is the standard error expressed as a percentage of the estimate.

The 95% confidence interval is the estimate + or - 2 times the standard error.

The program also calculates sampling errors as if the sample had been drawn as a simple random sample (SRS). The variance of a ratio r = y / x from a simple random sample is computed using the following formula:

$$\text{var}(r) = \frac{1-f}{n\bar{x}^2} \sum_{i=1}^{n} \frac{(y_i - rx_i)^2}{n-1}$$

where n is the unweighted number of observations

The sum of the squares above is calculated as:

$$\sum_{i=1}^{n} y_i^2 - 2r\sum_{i=1}^{n} x_i y_i + r^2 \sum_{i=1}^{n} x_i^2$$

The program accumulates the individual values of x, y, x^2, y^2 and xy as the database is being processed. When all the data have been processed, the SRS variance is calculated according to the formula above. SRS standard error and relative standard error are derived as described above.

Appendix B - ancillary table formats

SErrors - Sampling errors

Name	Description	Type
StatNo	Estimate sequence number	Integer 2
SGVariable	Subgroup variable number	Byte 1
	0 National	
	1 Urban/Rural	
	2 Regional	
SGNumber	Subgroup entry number	Byte 1
SGName	Subgroup entry name	Text 16
EstimateName	Estimate name	Text 12
Estimate	Value of the estimate	Double 8
EstimateN	Number of estimates (n-1 sub-samples)	Long 4
EstimateVar	Variance of the estimate	Double 8
EstimateSE	Standard error of the estimate	Double 8
RelSe	Relative standard error of the estimate	Double 8
CIMin	Minimum for 95% confidence interval	Double 8
CIMax	Maximum for 95% confidence interval	Double 8
NObs	Weighted number of observations	Double 8
Nunw	Unweighted number of observations	Long 4
Runw	Unweighted estimate value	Single 4
srsVar	Variance if simple random sample	Double 8
srsSE	Standard error if simple random sample	Single 4
srsRelSE	Relative standard error if simple random sample	Single 4

SErrorsW - Cluster work table

Name	Description	Type
A1	Cluster number	Text 3
SeqNo	Sequence number	Integer 2
EstimateName	Estimate name	Text 12
Y	Numerator total	Double 8
X	Denominator total	Double 8
UrbRur	Urban rural code	Text 1
Region	Region code	Text 2

Structure and logic of module: mCoreInd – calculate core indicators and sampling errors

Data organization:

Custom data types:

```
Type TotStats
  Id As Integer         ' Id for active entry = i or 0 if inactive
  r As Double           ' Value of the estimate
  y As Double           ' Numerator
  x As Double           ' Denominator
  RSum As Double        ' Sum of individual sample estimates
  VarSum As Double      ' Sum of individual sample components of variance
  n As Long             ' Number of observations (clusters)

' unweighted statistics for calculating SRS variance
  Nobs As Long          ' Number of observations (HH/Ind/Ch)
  Nunw As Long          ' Number of observations (HH/Ind/Ch)
  Xunw As Double        ' Sum of x(i)
  Yunw As Double        ' Sum of y(i)
  X2unw As Double       ' Sum of x(i) squared
  XYunw As Double       ' Sum of x(i) * y(i)
  Y2unw As Double       ' Sum of y(i) squared
  Runw As Single        ' Unweighted r (estimate)
  srsVar As Double      ' simple random sample variance
  srsSE As Single       ' simple random sample standard error
  srsRelSE As Single    ' simple random sample relative standard error
End Type

Type ClStats
  y As Double                   ' Numerator
  x As Double                   ' Denominator
  r As Double                   ' Value of estimate
  Nobs As Long                  ' Number of observations (HH/Ind/Ch)
  Nunw As Long                  ' Number of observations (HH/Ind/Ch)
  Xunw As Double                ' Sum of x(i)
  Yunw As Double                ' Sum of y(i)
  X2unw As Double               ' Sum of x(i) squared
  XYunw As Double               ' Sum of x(i) * y(i)
  Y2unw As Double               ' Sum of y(i) squared
End Type
```

Data definitions:

Accumulators for basic statisitics by indicator (Type is TotStats)

TS	Totals for level of aggregation
Rural	Totals for rural clusters
RuralPoor	Totals for poor households (Quintile = 1)
Urban	Totals for urban clusters
UrbanPoor	Totals for poor households (Quintile = 1)

Accumulators for basic statistics by cluster (Type is clStats)

CS	Cluster totals
CSpoor	Cluster totals for poor households (Quintile = 1)

Cluster level classification variables:

```
Dim Year As String              ' Year of survey
Dim Level As String             ' Level of aggregation
Dim GPZone As String            ' Geo-political Zone
Dim State As String             ' State code
Dim SenZone As String           ' State Senatorial Zone
Dim FedConst As String          ' State Federal constituency
Dim LGA As String               ' Local government area
Dim A1 As String                ' Cluster number
```

Cluster identification variables:

```
Dim ClusterId As String
Dim PrevClusterId As String
Dim clUrbRur As String
```

Procedure definitions:

dbInit

initStat	initialize accumulators for a single indicator
initStats	initialize accumulators for all indicators
initCluster	initialize cluster accumulators for all indicators
endCluster	add cluster totals to indicators, write cluster data to work table