SISTEMAS INTEGRALES
Leading Survey Design and Implementation Worldwide

Suriname High Frequency Survey (HFS) - 2022
Sampling Design and Weighting

Suriname High Frequency Survey (HFS) - 2022
Sampling Design and Weighting

The sample of the Suriname HFS-2022 was generated through a Random Digit Dialing (RDD) process covering all cell phone numbers active at the time of the sample selection.

Survey estimates represent households with at least one cell phone and individuals of 18 years of age or above who have an active cell phone number at home.

1. Sampling design

The RDD methodology generates virtually all *possible* cell phone numbers in the country under the national telephone numbering plan and then draws a random sample of numbers. This method guarantees full coverage of the population with a phone.[1]

First, a large first-phase sample was selected from the frame of numbers. This sample was then screened through an automated process to identify the active numbers. The active numbers were then cross-checked with business registries (based on yellow page directories and websites) to identify and remove business numbers not eligible for this survey.

A smaller second-phase sample[2] was then selected from the *active* numbers identified in the first-phase sample and was delivered to the operations team to be contacted by the interviewers.[3]

2. Weighting

This survey has two sample units: households and individuals. Sampling weights were computed for each unit and should be used according to the estimate of interest. The weighting process involves five steps:

1. Calculation of the inclusion probabilities of cell phone numbers.
2. Computation of design weights for households and individuals.
3. Weighting of data on children & adolescents.
4. Nonresponse weighting adjustment.
5. Calibration of individual and household weights, using external data from official sources (adjusted for the national phone coverage).

---

[1] Given that the survey used a sampling frame of telephone numbers, results represent the population with at least one active phone and exclude the population with no phone.

[2] Note that the selection of phone numbers involves two sampling *phases*, and not two sampling *stages*. The survey involves only one sampling stage.

[3] Furthermore, the second-phase sample was delivered in batches to the country teams during fieldwork. Delivering large lists of numbers could have facilitated the “misuse” of the sample by easily replacing non-answering numbers, raising nonresponse rates and potentially increasing nonresponse biases.

Step 1: Inclusion probabilities of cell phone numbers

A first-phase sample of numbers was selected with simple random selection without replacement from the entire frame. The selected numbers were then screened and classified into active and inactive.

The first-phase inclusion probabilities of cell phone numbers are[4]

$$p = \frac{n_{(1)}^{C}}{N_{(1)}^{C}} = \frac{n_{(1)A}^{C} + n_{(1)IN}^{C}}{N_{(1)}^{C}}$$

where

$p_{(1)i}^{C}$ is the first-phase inclusion probability of the i-th cell phone number;

$n_{(1)}^{C}$ is the size of the first-phase sample of cell phones, composed of $n_{(1)A}^{C}$ active cell phones and $n_{(1)IN}^{C}$ inactive cell phones;

$N_{(1)}^{C}$ is the cell phone frame size, the total number of all possible cell phones according to the national numbering plan;

Next, a second-phase sample was selected systematically out of the first-phase samples of active cell telephone numbers. The second-phase inclusion probabilities of cell phones are

$$p_{(2)i|(1)i}^{C} = \frac{n_{(2)A}^{C}}{n_{(1)A}^{C}}$$

where

$p_{(2)i|(1)i}^{C}$ is the second-phase inclusion probability of the i-th active cell phone number conditional on being selected in the first phase;

$n_{(2)A}^{C}$ is the size of the second-phase sample of active cell phones;

The unconditional inclusion probabilities of the second-phase active cell phones are

$$p_{i}^{C} = p_{(1)i}^{C}\, p_{(2)i|(1)i}^{C} = \frac{n_{(1)A}^{C} + n_{(1)IN}^{C}}{N_{(1)}^{C}} \frac{n_{(2)A}^{C}}{n_{(1)A}^{C}} = \frac{n_{(1)A}^{C} + n_{(1)IN}^{C}}{n_{(1)A}^{C}} \frac{n_{(2)A}^{C}}{N_{(1)}^{C}} = \frac{n_{(2)A}^{C}}{\widehat{RA}_{(1)}^{C} N_{(1)}^{C}} = \frac{n_{(2)A}^{C}}{\hat{A}_{(1)}^{C}}$$

[4] Inclusion probabilities of cell phones do not show a stratum index since most cell phone samples were not stratified for the reasons stated above.

where $\widehat{RA}_{(1)}$ is the rate of active phones estimated in the first phase.[5] Hence, the unconditional inclusion probabilities of the second-phase active numbers π_i^C can be expressed as the ratio between the active numbers selected in the second phase and an estimate of the total active numbers in the frame $\hat{A}_{(1)}$.

Step 2: Design weights for households and individuals

The selection probabilities of households and individuals aged 18 years and older are based on the inclusion probabilities of the cell phones through which they can be reached. Therefore, the computation of household and individual weights should account for multiple chances of selection. This multiplicity weighting adjusts estimates to eliminate the over-representation of households and individuals in the sample that can be reached through more telephone numbers than other households and individuals. It thus eliminates the chance for multiplicity sampling bias.

Multiplicity adjustment

There is multiplicity probability when a household has a larger selection probability because it can be selected through different sample elements (telephone numbers). Households with more than one cell phone number are over-represented in sample designs like this. As a result, their selection probabilities need to be adjusted to account for this increased chance of selection. The multiplicity-adjusted *household* selection probabilities are computed as

$$p_{mj}^C = m_{cj} p_i^C$$

where

p_{mj}^C is selection probability of the j-th household when contacted through a cell phone, adjusted for multiplicity of working cell phones in the household;

m_{cj} is the number of working cell phones in the j-th household;

Therefore, if a household has m_c cell phones, its chance of being selected through a cell phone is m_c higher than a household where there is only one cell phone. Since the number of cell phones in a household is unknown at the time of the sample design, it needs to be asked during the interview in the questionnaire. For this purpose, the survey collected information about the number of cell phones in the respondent households through the following question:

How many working cell phones in total are owned by the persons in your household, including you?

The probability of an *individual* being selected through a cell phone equals the inclusion probability of his or her cell phone number.

$$p_k^C = p_i^C$$

where

[5] $\widehat{RA}_{(1)}$ estimates are highly precise due to the very large size of the first-phase samples.

p_k^C is the selection probability of the k-th individual when contacted through a cell phone.

Household and individual design weights, w_{0j} and w_{0k} respectively, are the inverse of the above selection probabilities

$$w_{0j} = p_j^{-1}$$

$$w_{0k} = p_k^{-1}$$

Step 3: Weighting of data on children & adolescents

HFPS collected specific data about a randomly selected child or adolescent 6 through 17 years of age in each interviewed household and also collected information for a randomly selected child between 0 and 5 years of age. To implement this, the questionnaire first collected a roster of all children and adolescents living in each respondent household and selected one at random.

The child/adolescent weight is based on his/her probability of selection within the household, conditional on his/her household being selected in the sample. Hence

$$p_{nj}^C = p_j^C / \sum_j n$$

where

p_{nj}^C	is the selection probability of the n-th child/adolescent in the j-th household when the household is contacted through a cell phone
p_j^C	is selection probability of the j-th household when contacted through a cell phone, adjusted for multiplicity of working cell phones in the household;
$\sum_j n$	is the number of eligible children and adolescents (6-17 years old) in the j-th household
$w_{0n} = 1/p_{nj}^C$	Children and adolescents' design weight w_{0n} is the inverse of the above selection probabilities

Step 4: Nonresponse adjustment

When a phone number is called, it is not always possible to carry out an interview. Nonresponse occurs because of a number of constraints. Most common are that nobody answers the call (no contact), the respondent is unwilling to cooperate (refusal), or language barriers exist.

The design weights of responding households and individuals were adjusted for nonresponse. This adjustment is based on the inverse of the weighted response rate estimate. This is the ratio of the sum of the design weights of all units (respondents and nonrespondents) to the sum of the design weights of respondents.

$$a_j = \frac{\sum_{j,R} w_{0j} + \sum_{j,NR} w_{0j}}{\sum_{j,R} w_{0j}} \quad ; \quad a_k = \frac{\sum_{k,R} w_{0k} + \sum_{k,NR} w_{0k}}{\sum_{k,R} w_{0k}}$$

where a_j is the nonresponse adjustment factor that should be applied to responding households and a_k is the nonresponse adjustment factor for responding individuals. R and NR indicate the responding and nonresponding units, respectively.

Thus, the nonresponse adjusted weights for responding households and individuals are

$$w'_j = w_{0j}\, a_j \quad ; \quad w'_k = w_{0k}\, a_k$$

Step 5: Calibration of individual and household weights

Finally, the weights for the responding households and individuals were calibrated to reflect the total population with phone by sex, age, and region available from external national official sources.

Calibration works by minimizing a measure of the distance between the input weights (nonresponse adjusted weights in this case) and the calibrated weights, under the constraint that the sum of the calibrated weights equals the sum of the totals of the auxiliaries from the external source. Unlike the nonresponse adjustment, weights calibration requires auxiliary variables for respondents only.

Among the existing calibration techniques, this survey applied the raking method, using the logit distance function. This method was most suitable given that all available auxiliary variables (region, sex, and age groups) were categorical, the region variable had many categories and the sample was rather small.

The final weights for responding households and individuals can then be expressed as

$$w_j = w'_j\, g_j = w_{0j}\, a_j\, g_j$$

$$w_k = w'_k\, g_k = w_{0k}\, a_k\, g_k$$

where

w_{0j} is the design weight for the j-th household;

a_j is the nonresponse adjustment factor for households;

g_j is the calibration factor for the j-th household;

w_{0k} is the design weight for the k-th individual;

a_k is the nonresponse adjustment factor for individuals; and

g_k is the calibration factor for the k-th individual.