Sampling weights – a Brief Description[1]

The COVID-19 Armenia High Frequency Survey (COVID-19 AHFS) was conducted from July to September 2020 by the UNICEF and the World Bank's Poverty and Equity team in collaboration with the R-Insights Research Company. The COVID-19 AHFS comprises two parts – survey I and II. In survey I, 4,555 households (responded either by the head of the household or someone else) were surveyed, out of which 2,653 households were eligible for the follow up interview in survey II.[2] Out of those eligible households, 1,648 were interviewed in survey II, which focused on the issues about children 0-18 years old.[3]

The samples are drawn from across the country, however, there was no weight distribution for the samples. In the raw data, Yerevan and other urban areas were undersampled and rural areas were oversampled. However, the base weights were 1 as it was a pure random digit dialing with no stratification.

Sampling weights are adjustment factors applied to each observation during the analysis to adjust for differences in the probability of selection between cases in a sample, either due to design or happenstance. The ILCS 2018 of Armenia was used to adjust weights for the respondents for the Armenia High Frequency survey. A propensity score model was applied, and the sample was stratified by marz and urban/rural.

To adjust the weights, the naive frequency weights (equals 1 for everybody) were first adjusted by marz (province) and urban/rural total counts of households using the ILCS 2018 numbers. The next step was to develop a propensity score model by searching for the overlapping variables in the HFS survey and ILCS survey at the household level. We were able to find the following variables: the household's ownership of TV, at least one computer, has at least one smartphone, and accessibility to an internet connection.

Next, logistic regression was run by setting dependent variable equals to 1 for the households in the HFS survey and equals to 0 if in the ILCS survey. We then used the predicted values to adjust the HFS weight further by taking the inverse of the predicted values (= 1/predicted values). Further, we refined the weights by collapsing the weights to decile and multiply by the pooled weights.

We have the following steps

- HFS: weight_1 = 1
- HFS: weight_2 = ILCS weight adjusted by marz-Urban/Rural
- HFS: weight_3 = weight_2 * continuous ps adjustment
- HFS: weight_4 = weight_3 * decile (ps adjustment)

---

[2] Households that have children younger than 18 years were eligible.
[3] An attrition rate was 38%.

Thus far, we have only adjusted the proportional weights at the household level. There could be a distortion of respondent rates at an individual level. To correct for this distortion, we use raking weights for individual-level characteristics which will reduce the distortions generated by the mobile phone instrument.

The individual-level characteristic used is: age group (18-25, 26-35, 36-45, 46-55, 56-65 and 65+), female dummy, educational level (4 categories – no education, primary, secondary and tertiary), household size, and marz-urban\level stratification. The command *maxentropy* in Stata is used. The command assigns a final weight between 0-1 for each individual, to which we multiply by the number of total population 18 and above from ILCS 2018 to get final weights for wave I data.

For survey 2, we follow the similar method above. First, instead of all households, we calculate the total number of households with children younger than 18 years old by marz urban/rural. Next, we applied the raking method to adjust for individual weights using female dummy, age groups (0-2, 2-6 and 6-18), household size, number of children under 18 years old and marz urban/rural stratification. The final weight was multiplied by the total number of population 0-18 years. The following table shows the distribution of samples by marz and urban/rural before and after the weight adjustments.

Table 1: Percentages of samples by marz with and without weight adjustments

Marz	Survey I (Unit: Household)			ILCS 2018	Survey II (Unit: HH with Children younger than 18)			ILCS 2018
	N	Percent (without weight)	Percent (Weighted)	Percent (Weighted)	N	Percent (without weight)	Percent (Weighted)	Percent (Weighted)
Yerevan	684	15.01%	33.90%	33.75%	200	4.39%	30.72%	28.44%
Aragatsotn	384	8.43%	4.22%	4.33%	143	3.14%	4.32%	4.58%
Ararat	386	8.47%	7.65%	8.42%	133	2.92%	10.03%	9.95%
Armavir	385	8.45%	9.09%	8.12%	145	3.18%	10.00%	11.03%
Gegharkunik	389	8.54%	6.08%	6.81%	176	3.86%	4.46%	5.19%
Lori	390	8.56%	9.64%	10.25%	152	3.34%	8.63%	8.90%
Kotayk	389	8.54%	9.28%	8.86%	146	3.20%	12.49%	11.75%
Shirak	387	8.49%	9.15%	8.40%	138	3.03%	9.11%	9.60%
Sjunik	386	8.47%	4.46%	4.74%	143	3.14%	3.28%	3.24%
VayotsDzor	385	8.45%	1.84%	1.79%	129	2.83%	1.81%	1.92%
Tavush	391	8.58%	4.70%	4.54%	143	3.14%	5.17%	5.40%
Urban/Rural					**Urban/Rural**			
Yerevan	684	15.01%	33.90%	33.75%	200	4.39%	30.72%	28.44%
Other Urban	1665	36.57%	31.20%	31.40%	580	12.73%	30.82%	30.07%
Rural	2206	48.42%	34.90%	34.85%	868	19.05%	38.46%	41.49%