Sample Weights

Sample weights are necessary to make accurate estimates in the population from a complex household survey. Sample weights are presented below. Inputs include three sources of information:

- GridSample: PSU population and strata population
- Fieldwork: Number of segments created in the PSU during pre-field enumeration, number of segments created in the PSU during post-field enumeration (one-stage only), total number of dwellings/households listed in the final (segmented) PSU, and number of households selected for sampling (two-stage only)
- Interview: Number of non-responding households per PSU (household questionnaire), and number of non-responding individuals per PSU (individual questionnaire)

Household sample (base) weight. The formulas use 2 indices: $1…k$ strata (or entire coverage area) and $1…i$ PSU. The household sample (design) weight – the probability that PSU i is selected – is given by:

$$w_{hh.b} = \frac{G_k/g_{ik}}{n_k} \times \frac{M_{ik}}{m_{ik}} \times (b1_{ik} * b2_{ik})$$

Where:

n_k is the number of selected PSUs in stratum k

G_k is the estimated total population in stratum k from GridSample

g_{ik} is the estimated population in PSU i in stratum k from GridSample

m_{ik} is the number of households sampled in PSU i and stratum k during fieldwork

M_{ik} is the number of total households enumerated in PSU i and stratum k during fieldwork

$b1_{ik}$ is the number of pre-field segments

$b2_{ik}$ is the number of post-field segments

Household response weight. Interviewers will list households and record household and individual response rates during fieldwork. After interviews are completed, calculate household response weight - the probability that cluster i is found and sampled, and households are found and respond – is given by:

$$w_{hh.r} = \frac{n_k}{n_{k*}} \times \frac{m_k}{m_{k*}}$$

Where:

> n_k is the number of selected PSUs in stratum k
> m_k is the number of households sampled in stratum k during fieldwork
> n_{k*} is the number of found and sampled PSUs in stratum k
> m_{k*} is the number of found and responded households in stratum k

Household sample weight. To calculate the raw household sample weight, multiply the sample base weight and household response weight like this:

$$w_{hh} = \frac{G_k/g_{ik}}{n_k} \times \frac{M_{ik}}{m_{ik}} \times (b1_{ik} * b2_{ik}) \times \frac{n_k}{n_{k*}} \times \frac{m_k}{m_{k*}}$$

Note that in one-stage samples, $\frac{M_{ik}}{m_{ik}}$ is equal to 1, and in two-stage samples, $b2_{ik}$ is equal to 1.

Individual sample weight. The individual sample weight includes four additional terms to account for the sampling of one adult among all eligible adults in the household, and the response rate of those adults. The individual sample weight is given by:

$$w_{ind.s} = \frac{G_k/g_{ik}}{n_k} \times \frac{M_{ik}}{m_{ik}} \times (b1_{ik} * b2_{ik}) \times \frac{n_k}{n_{k*}} \times \frac{m_k}{m_{k*}} \times \frac{U_{ik}}{u_{ik}} \times \frac{u_k}{u_{k*}}$$

Where:

> n_k is the number of selected PSUs in stratum k
> G_k is the estimated total population in stratum k from GridSample
> g_{ik} is the estimated population in PSU i in stratum k from GridSample
> m_{ik} is the number of households sampled in PSU i and stratum k
> M_{ik} is the number of total households enumerated in PSU i and stratum k
> $b1_{ik}$ is the number of pre-field segments
> $b2_{ik}$ is the number of post-field segments
> n_{k*} is the number of found and sampled PSUs in stratum k
> m_k is the number of households sampled in stratum k during fieldwork
> m_{k*} is the number of found and responded households in stratum k
> U_{ik} is the number of eligible adults in PSU i and stratum k
> u_{ik} is the number of sampled adults in PSU i and stratum k
> u_k is the number of sampled adults in stratum k
> u_{k*} is the number of responded adults in stratum k

Calculate sample weights in Excel. Calculate sample weights in Excel, or a statistical software programme such as Stata. See Figure 25 for example.

Normalizing sample weights. Household surveys are often "normalized" or "standardized" such that the sum of weighted respondents equals the sum of respondents. Conceptually, each observation in the sample represents slightly more or slightly less than 1 household or person. To normalize sample weights, apply the below formulas:

$$w_{hh\_norm} = w_{hh} \times \frac{\sum(m_{ik*})}{\sum(w_{hh} \times m_{ik*})}$$

Where:

m_{ik*} is the number of adults with a completed interview in PSU *i* in stratum *k*
w_{hh} is the raw household sample weight in PSU *i* in stratum *k*

$$w_{ind\_norm} = w_{ind} \times \frac{\sum(u_{ik*})}{\sum(w_{ind} \times u_{ik*})}$$

Where:

u_{ik*} is the number of adults with a completed interview in PSU *i* in stratum *k*
w_{ind} is the raw individual sample weight in PSU *i* in stratum *k*

PSU_ID	Stratum_ID	g_ik	n_k	G_k	1/b_ik	M_ik	m_ik	n_k*	m_k	m_k*	m_ik*	w_hh	w_hh * m_ik*	w_hh_norm
1	1	872.0	15	112020.6	0.071	12	12	15	292	292	12	119.901	1438.809	0.930
2	2	91.4	5	1318.6	0.059	11	11	5	95	90	11	51.763	569.392	0.401
3	1	830.1	15	112020.6	0.063	24	24	15	292	292	24	143.936	3454.476	1.116
4	1	245.5	15	112020.6	0.091	24	24	15	292	292	24	334.641	8031.387	2.595
6	1	888.2	15	112020.6	0.063	16	16	15	292	292	16	134.527	2152.435	1.043
7	2	230.7	5	1318.6	0.071	27	27	5	95	90	26	16.890	439.150	0.131
8	2	94.1	5	1318.6	0.024	12	12	5	95	90	12	124.305	1491.654	0.964
9	1	1142.9	15	112020.6	0.067	25	25	15	292	292	25	98.016	2450.404	0.760
10	1	317.1	15	112020.6	0.077	16	16	15	292	292	16	306.135	4898.162	2.374
12	1	702.8	15	112020.6	0.083	18	18	15	292	292	18	127.509	2295.163	0.989
13	1	617.3	15	112020.6	0.071	7	7	15	292	292	7	169.376	1185.629	1.314
14	2	172.7	5	1318.6	0.091	15	15	5	95	90	15	17.735	266.018	0.138
15	2	133.7	5	1318.6	0.143	30	30	5	95	90	26	14.573	378.888	0.113
16	1	721.4	15	112020.6	0.053	21	21	15	292	292	21	196.692	4130.526	1.525
18	1	728.6	15	112020.6	0.111	25	25	15	292	292	25	92.244	2306.104	0.715
20	1	844.2	15	112020.6	0.059	19	19	15	292	292	19	150.396	2857.520	1.166
22	1	871.0	15	112020.6	0.067	12	12	15	292	292	12	128.607	1543.286	0.997
23	1	840.8	15	112020.6	0.063	15	15	15	292	292	15	142.106	2131.588	1.102
24	1	605.1	15	112020.6	0.111	35	35	15	292	292	35	111.083	3887.898	0.861
25	1	666.7	15	112020.6	0.077	23	23	15	292	292	23	145.617	3349.183	1.129
TOTAL						387	387				382	2626.051	49257.672	
											Σm_ik^*		$\Sigma(w\_hh * m\_ik^*)$	

PSU_ID	Stratum_ID	w_hh	u_k	U_ik	u_ik	u_ik*	u_k*	w_ind	w_ind * u_ik*	ind_norm
1	1	119.901	292	34	12	12	288	344.437	4133.245	0.979
2	2	51.763	90	35	11	11	88	168.443	1852.878	0.479
3	1	143.936	292	67	24	23	288	407.404	9370.283	1.158
4	1	334.641	292	62	24	24	288	876.496	21035.913	2.491
6	1	134.527	292	32	16	16	288	272.791	4364.659	0.775
7	2	16.890	90	84	26	26	88	55.809	1451.037	0.159
8	2	124.305	90	30	12	12	88	317.824	3813.889	0.903
9	1	98.016	292	63	25	25	288	250.431	6260.782	0.712
10	1	306.135	292	40	16	15	288	775.968	11639.513	2.205
12	1	127.509	292	49	18	18	288	351.929	6334.722	1.000
13	1	169.376	292	19	7	7	288	466.119	3262.832	1.325
14	2	17.735	90	39	15	14	88	47.158	660.208	0.134
15	2	14.573	90	54	26	25	88	30.954	773.852	0.088
16	1	196.692	292	52	21	21	288	493.811	10370.024	1.403
18	1	92.244	292	86	25	25	288	321.727	8043.178	0.914
20	1	150.396	292	46	19	19	288	369.173	7014.293	1.049
22	1	128.607	292	27	12	12	288	293.385	3520.620	0.834
23	1	142.106	292	43	15	15	288	413.028	6195.421	1.174
24	1	111.083	292	136	35	34	288	437.631	14879.453	1.244
25	1	145.617	292	52	23	22	288	333.793	7343.441	0.949
TOTAL				1050	382	376	4760	7028.311	132320.242	
						Σu_ik^*			$\Sigma(w\_ind * u\_ik^*)$	

Figure 26. Example household and individual sample weights calculations