STAT 403/603 Tutorial - Week 9.

This tutorial includes a review of concepts such as sampling designs, inclusion probabilities and Horvitz-Thompson estimator.

Example. Suppose that we are interested in the total number of palm trees on three islands. Due to some practical reasons such as budget and resources, we are only allowed to investigate the exact number of palm trees on at most two islands. Therefore, we decide to use a simple random sampling with replacement of size 2. That is, our sample is obtained by two independent draws from $\mathcal{U} = \{1,2,3\}$, with the following selection probabilities (note this is not inclusion probabilities) at each draw:

	unit (islands)	1	2	3
(i)				
(p_i)	Prob.	0.3	0.2	0.5
(y_i)	# trees	11	6	25

Describe the sampling design, inclusion probabilities for each unit and the Horvitz-Thompson estimator of the population total (total number of palm trees) for every possible sample.

Solution. There are $2^3 = 8$ ~~possible~~ samples (subsets of $\mathcal{U}$). The sampling design is the probability distribution over these samples:

Sample	ϕ	$\{1\}$	$\{2\}$	$\{3\}$	$\{1,2\}$	$\{1,3\}$	$\{2,3\}$	$\{1,2,3\}$
Prob.	0	0.09	0.04	0.25	0.12	0.3	0.2	0

e.g.

$P(\{1\}) = P(\text{first draw} = 1, \text{ second draw} = 1) = 0.3 \times 0.3 = 0.09$;

$P(\{1,2\}) = P(\text{first draw} = 1, \text{ second draw} = 2) + P(\text{first draw} = 2, \text{ second draw} = 1)$

$= 0.3 \times 0.2 + 0.2 \times 0.3 = 0.12$

Using the sampling design, we can easily figure out the inclusion probabilities π_i for each $i \in \mathcal{U}$:

unit i	1	2	3
π_i	0.51	0.36	0.75

e.g. $\pi_1 = P(\{1\}) + P(\{1,2\}) + P(\{1,3\})$

$= 0.09 + 0.12 + 0.3 = 0.51$.

Now we're able to calculate the Horvitz-Thompson estimates $\hat{T} = \sum_{i \in u} \frac{y_i}{z_i}$ of the total palm trees for every possible sample.

sample	corresponding y_i's	HT estimates
$\{1\}$	$y_1 = 11$	$y_1/z_1 = 11/0.51 = 21.57$
$\{2\}$	$y_2 = 6$	$y_2/z_2 = 6/0.36 = 16.67$
$\{3\}$	$y_3 = 25$	$y_3/z_3 = 25/0.75 = 33.33$
$\{1,2\}$	$y_1 = 11,\ y_2 = 6$	$y_1/z_1 + y_2/z_2 = 11/0.51 + 6/0.36 = 38.24$
$\{1,3\}$	$y_1 = 11,\ y_3 = 25$	$y_1/z_1 + y_3/z_3 = 11/0.51 + 25/0.75 = 54.9$
$\{2,3\}$	$y_2 = 6,\ y_3 = 25$	$y_2/z_2 + y_3/z_3 = 6/0.36 + 25/0.75 = 50$

You can easily verify that H-T estimator is unbiased.