Post-Stratification

May 2022

1. Sampling theory

- $U = \{1, \cdots, N\}$, the elements of U are called labels or identifiers. A set of identifiers is a *sample*
- *Sampling frame*, a set associated 1-1 with a subset of U
- $\mathcal{A} \subset \mathcal{P}(U)$, the set of possible samples under a certain probabilistic procedure.
- Associated to U is $\mathcal{F} = \{y_i\}$. It is denoted by *finite population* or *finite universe*.
- If y_i's are vector valued, it is called vector of characteristics.

Let P a probability measure on $\mathcal{A}$. A desing is a map $p : \mathcal{A} \to [0, 1]$ such as $p(s) = P[s \in \mathcal{A}]$. It represent the probability the sample are selected by the sampling process.

Let $I_i : \mathcal{A} \to \{0, 1\}$ the indicator variable for element i , i.e. $I_i(s) = 1$ if i is in s and zero otherwise.

The sampling desing determines the probabilistic structer under $\mathbf{d} = (I_1, \cdots I_N)$.

In the context of non replacement sampling, we define the initial probability as

$$\pi_i = P(i \in s) = \sum_{a \in s(i)} p(a)$$

where $s(i)$ is the set of samples containing i. We note that $\pi_i = E[I_i]$.

The sample sum is

$$\sum_{i=1}^{N} I_i(s) y_i = \sum_{i \in s} y_i$$

Definition 1 *A statistic $\hat{\theta}$ is desing unbiased for $\theta_N = (\theta_1, \cdots, \theta_N)$ if*

$$E[\hat{\theta}|\mathcal{F}] = \theta_N = (\theta_1, \cdots, \theta_N)$$

for any vector $(y_1, \cdots, y_N)$.

The expectation conditional to the population parameters is often referred as expectation desing. If the y_i's are random, we can combine the estimation of the characteristics with the sample desing. It is called a *strategy*.

Two population parameters of special iterest are the total

$$T_y = \sum_{i=1}^{N} y_i \tag{1}$$

and the mean $\overline{y}_N = N^{-1} T_y$

Referencias

[Cochran] Cochran, W. G. (1997), Sampling Techniques, John Wiley & Sons

[Füller] Füller, W. A. (2009), Sampling Statistics, John Wiley & Sons

[Ian Laga et al.] Ian Laga, Le Bao, Xiaoyue Niu. (2020/21) Thirty Years of The Network Scale up Method

[Appendix] Ian Laga, Le Bao, Xiaoyue Niu. (2020/21) Thirty Years of The Network Scale up Method Appendix

[Habecker et al.] Habecker, P., Dombrowski, K., and Khan, B. (2015). Improving the network scale-up esti- mator: Incorporating means of sums, recursive back estimation, and sampling weights. PloS one, 10(12).

[Killworth et al. (a)] Killworth, P. D., Johnsen, E. C., McCarty, C., Shelley, G. A., and Bernard, H. R. (1998a). A social network approach to estimating seroprevalence in the united states. Social networks, 20(1):23?50.

[Killworth et al. (b)] Killworth, P. D., McCarty, C., Bernard, H. R., Shelley, G. A., and Johnsen, E. C. (1998b).

Estimation of seroprevalence, rape, and homelessness in the united states using a social network approach. Evaluation review, 22(2):289?308

[D. Holt et al.] D. Holt and T. M. F. Smith (1979). Journal of the Royal Statistical Society. Series A (General), Vol. 142, No. 1,pp. 33-46

[Ehsan Zamanzadea et al.] Ehsan Zamanzadea, Xinlei Wangb (). Estimation of population proportion for judgment post-stratification

[Feehan and Salganik] Feehan, D.M. and Salganik, M. J. (2016). Generalizing the network scale-up method: a new estimator for the size of hidden populations. Sociological methodology, 46(1):153?186.