City University of New York (CUNY)

CUNY Academic Works

Publications and Research

Hunter College

1962

ON TWO METHODS OF UNBIASED ESTIMATION WITH AUXILIARY VARIATES

William (Bill) H. Williams
CUNY Hunter College

[How does access to this work benefit you? Let us know!](#)

More information about this work at: https://academicworks.cuny.edu/hc_pubs/662

Discover additional works at: <https://academicworks.cuny.edu>

This work is made publicly available by the City University of New York (CUNY).
Contact: AcademicWorks@cuny.edu

ON TWO METHODS OF UNBIASED ESTIMATION
WITH AUXILIARY VARIATES

BY

W. H. WILLIAMS

Reprinted from the JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
March 1962, Vol. 57
pp. 184-186

ON TWO METHODS OF UNBIASED ESTIMATION WITH AUXILIARY VARIATES

W. H. WILLIAMS

Bell Telephone Laboratories, Incorporated

This paper shows the relationship between two existing methods of generating unbiased ratio and regression estimators. One procedure involves splitting up the population (theoretically) while the other splits up the sample itself. The two procedures yield similar estimators and are combinatorially equivalent. They are unlike in that they utilize different conditional arguments.

1. INTRODUCTION

IN A paper by Williams [6], a procedure for generating unbiased estimators was introduced. A method was also given by Mickey [4]. It is shown that a modification of Mickey's argument gives rise to estimators which have the same form as those presented by Williams. This paper deals with their equivalence in simple random sampling.

2. TWO METHODS OF GENERATING UNBIASED ESTIMATORS AND THEIR RELATIONSHIP

The problem arises in sampling theory with the use of ratio and regression procedures for estimating a population mean μ_Y aided by an auxiliary variate x . Analytically, we have a random sample of n pairs (y_i, x_i) from a population of size N . The object is to estimate μ_Y given that μ_X , the population x mean, is known exactly.

Mickey generates unbiased estimators by noting that $\bar{y} - a(\bar{x} - \mu_X)$ is an unbiased estimator of μ_Y for all choices of the constant a , where $\bar{y}$ and $\bar{x}$ are the sample means of the n pairs drawn without replacement from the population of size N . He then modifies this idea so that the coefficient can be a function of the observations while the estimator remains unbiased. To do this he considers the sample as an ordered set of elements and points out that for any choice of the first α sample elements, the $n - \alpha$ remaining sample elements constitute a random sample from the finite population of $N - \alpha$ elements. This method can then be used to construct various unbiased estimators of μ_Y , for it can be seen that

$$E_{\alpha} \left\{ \frac{n\bar{y} - \alpha\bar{y}(\alpha)}{n - \alpha} - a(z_{\alpha}) \left[\frac{n\bar{x} - \alpha\bar{x}(\alpha)}{n - \alpha} - \frac{N\mu_X - \alpha\bar{x}(\alpha)}{N - \alpha} \right] \right\} = \frac{N\mu_Y - \alpha\bar{y}(\alpha)}{N - \alpha} \quad (1)$$

where z_{α} denotes the ordered set of observations on the first α sample elements, and the order is, say the order in which the sample is drawn. Here $a(z_{\alpha})$ is a function of the first α sample elements and $\bar{y}(\alpha)$ and $\bar{x}(\alpha)$ are the sample means of the first α sample observations. E_{α} denotes the conditional expectation with respect to the first α sample elements.

By a simple rearrangement of Equation (1), it is seen that $E_{\alpha} \mu(\alpha) = \mu_Y$, where

$$t(\alpha) = \frac{N - \alpha}{N} \cdot \frac{n}{n - \alpha} \{ \bar{y} - a(z_\alpha) [\bar{x} - \mu_x] \} \\ - \frac{N - n}{N} \cdot \frac{\alpha}{n - \alpha} \{ \bar{y}(\alpha) - a(z_\alpha) [\bar{x}(\alpha) - \mu_x] \}. \quad (2)$$

It is a corollary of this that $t(\alpha)$ is an unconditionally unbiased estimate of μ_Y .

It is possible by different choices of the form of the $a(z_\alpha)$ coefficients to generate many different estimators. Several are considered by Mickey and the reader is referred to his paper for the details and numerical examples.

The procedure used by Williams [6] has the following stages. First, select randomly one of all possible splits of the population into s mutually exclusive groups of size n/k . Then $N = sn/k$.¹ At the second stage, select with equal probability and without replacement k of the s groups. Then the sample is of size n .

Next consider the conditional distribution for a fixed set of s groups. Attached to the s groups are characteristics $\bar{y}_{(i)}$, $\bar{x}_{(i)}$ and $b_{(i)}$, $(i) = 1, 2, \dots, s$, where $\bar{y}_{(i)}$ and $\bar{x}_{(i)}$ are means of the n/k units in the group and $b_{(i)}$ is an unspecified function of the y and x of that group. For a given split and a random selection of groups, the expectations of $\bar{y}_i$ and $\bar{x}_i$, $i = 1, 2, \dots, k$, are μ_Y and μ_X respectively. Furthermore,

$$\left(1 - \frac{k}{s}\right) \frac{1}{k(k-1)} \sum_{i=1}^k (b_i - \bar{b})(\bar{x}_i - \bar{x}) \quad (3)$$

is an unbiased estimator of $\text{Cov}(\bar{b}, \bar{x})$ where $\bar{b} = \sum_{i=1}^k b_i/k$.

Hence, by considering $\bar{y} + \bar{b}(\mu_X - \bar{x})$ and its expectation of $\mu_Y - \text{Cov}(\bar{b}, \bar{x})$, it is seen that

$$T_k = \bar{y} + \bar{b}(\mu_X - \bar{x}) + \left(1 - \frac{n}{N}\right) \frac{1}{k(k-1)} \sum_{i=1}^k (b_i - \bar{b})(\bar{x}_i - \bar{x}) \quad (4)$$

is a conditionally unbiased estimator of μ_Y . It is then unbiased unconditionally. This approach is valid for any defined form of the coefficients b_i . Numerical examples are presented by Williams [6].

A change in Mickey's argument is obtained by splitting the sample into k groups of equal size and then considering each in turn as the conditional group. This gives k estimators of the form $t_i(n/k)$, $i = 1, 2, \dots, k$ where

$$t_i\left(\frac{n}{k}\right) = \frac{1}{k-1} \cdot \frac{kN - n}{N} \{ \bar{y} + b_i(\mu_X - \bar{x}) \} \\ - \frac{1}{k-1} \cdot \frac{N - n}{N} \{ \bar{y}_i - b_i(\bar{x}_i - \mu_X) \} \quad i = 1, 2, \dots, k. \quad (5)$$

The subscript i denotes the group that is being used to form the coefficients, and the previous coefficient notation, $a(z_\alpha)$, (Equations (1) and (2)), has been simplified to b_i to correspond to the notation used for the arbitrary coefficient of Equations (3) and (4).

¹ It is assumed that this relationship holds in terms of integers.

Now if these k estimators are averaged (yielding an unbiased estimator of μ_Y), it will be of the form

$$T_k = \bar{y} + \delta(\mu_x - \bar{x}) + \left(1 - \frac{n}{N}\right) \frac{1}{k-1} \sum_{i=1}^k \frac{b_i - \bar{b}}{k} \frac{\bar{x}_i - \bar{x}}{k} \quad (6)$$

where

$$\delta = \sum_{i=1}^k b_i/k.$$

The average displayed in Equation (6) has been denoted T_k since Equations (4) and (6) are equivalent. Thus the two procedures yield estimators of the same form and they are combinatorially equivalent. In either case at most $N!/(N-n)!(n/k!)^k$ different values of the estimators are possible and each value is equally likely.

Both procedures generalize easily to p auxiliary variates.

3. FURTHER REMARKS

The purpose of splitting the population [6] is to generate unbiased estimators. The operational aspects of sampling must usually be to split the sample and not the population; then the T_k fall into Mickey's class of estimators. The interesting aspect is that to display unbiasedness the two procedures depend upon different conditional arguments. Furthermore, it is convenient to split the population into *equal* size groups such that $N = sn/k$. In this case the groups would be weighted equally, and so unequal probability sampling has not been considered as in such work as Des Raj [2] and Murthy [5].

REFERENCES

- [1] Cochran, W. G., *Sampling Techniques*. New York: John Wiley and Sons, 1953.
- [2] Des Raj, "Some Estimators in Sampling with Varying Probabilities without Replacement," *Journal of the American Statistical Association*, 51 (1956), 269-84.
- [3] Hartley, H. O., and Ross, A., "Unbiased Ratio Estimators," *Nature*, 174 (1954), 270-1.
- [4] Mickey, M. R., "Some Finite Population Unbiased Ratio and Regression Estimators," *Journal of the American Statistical Association*, 54 (1959), 594-612.
- [5] Murthy, M. N., "Ordered and Unordered Estimators in Sampling without Replacement," *Sankhyā*, 18 (1957), 379-90.
- [6] Williams, W. H., "Generating Unbiased Ratio and Regression Estimators," *Biometrics*, 17 (1961), 267-74.