Resource Inventory Notes

BLM 18 February 1979

MULTI-STAGE AND MULTI-PHASE SAMPLING

by

J.D. Nichols 1/

Historically sampling and estimation procedures have been misnamed and misused. One of the most abused of these terms is multi-stage sampling. As a result of several studies multi-stage sampling has taken on the connotation of any sampling and estimation procedure that calls on multiple levels of data to provide information.

Multi-stage sampling has a rather simple meaning.2/ The multi-stage sampling frame, Figure 1, consists of a list of sample units (Primary Sample Units)which are made up of smaller units (Secondary Sample Units), which are in turn made up of smaller units (Third Stage Sample Units). The number of stages and the size and number of sample units at each stage is dependent on the objectives of the inventory and the characteristics of the population to be sampled. The size and number of units may vary at each stage. The method of selecting a PSU, SSU and TSU for measurement depends on the objectives of the inventory and population characteristics. They can include simple random, variable probability and stratified selection.

The objective of multi-stage sampling is to provide a sample unit that can be cost efficiently measured and to cluster these samples in the higher level (larger) sample unit to reduce the travel cost between measurement units. The estimators for multi-stage sampling are well documented in several texts (see references) including Forest Mensuration, by B. Husch, C. Miller, and T. Beers.

1/ Owner, Resource Inventory Services, 650 Aram Ave, San Jose, Ca.

2/ Source: B. Husch, C. Miller, and T. Beers, Forest Mensuration, 2nd Edition, pages 219, 224, and 226.

Published by:
USDI, Bureau of Land Management, D 340
Denver Service Center, Denver Federal Center, Bldg. 50
Denver, Colorado, 80225

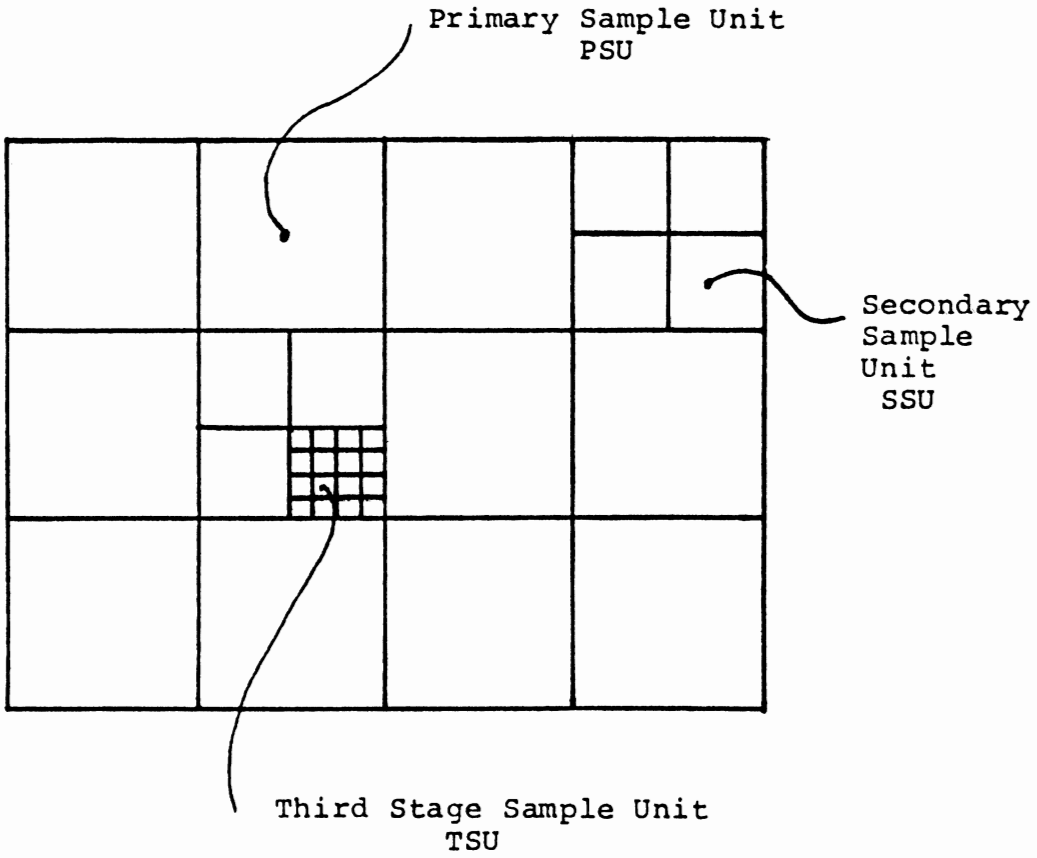

Figure 1: Hypothetical 3-stage multi-stage sample layout with equal size at each stage.

The optimization of a multi-stage design depends on the data collection cost, travel cost, between unit variability, and precision objectives. Given a two stage sampling frame the number of second stage samples is obtained from

$$n = \sqrt{\frac{C_1 S_W^2}{C_2 S_B^2}}$$

n = optimum number of secondary units per primary unit.

C_1 = cost of establishing a primary unit.

C_2 = cost of establishing and measuring a secondary unit.

S_W^2 = estimate of variance within the groups of secondary sample units.

S_B^2 = estimate of the variance between the means of the secondary sampling units within the primaries.

Given a fixed budget, the number of primaries is

$$m = \frac{C}{C_1 + nC_2}$$

The points to be made are: 1)the optimization is dependent on cost and variance and 2)the procedure does not use multiple levels of data.

The sampling scheme that is most often confused and confounded with multi-stage sampling is multi-phase sampling (double sampling for two phases). Multi-phase sampling in general consists of selecting a large first phase sample of an auxiliary variable (x_i) which will provide a precise estimate of the primary variable (y_i). A subset of the first phase units are then selected and accurate measurements of the primary variable (y_i) are taken. In operational forest inventories this has included basal area (x_i) as an auxiliary estimator of timber volume (y_i) and aerial photo estimates of timber volume (x_i) as an auxiliary variable for ground measured timber volume (y_i). In range sampling a common practice is to make an ocular (professional) estimate of the weight of available forage (x_i) as an auxiliary variable for clipped, oven dry weight of the forage (y_i) on sample plots.

The procedure assumes that the observation of the x_is are paired (Figure 2) with the y_is, that there is a strong relationship between the x_i and the y_i, that the relationship can be estimated and that the x_i measurement is much less expensive to collect and process per sample than the y_i measurement.

The procedure to estimate the relationship between the y and x can take one of several forms. The form depends on the functional relationship between the x and y observation and the form of distribution of errors in estimating the y from x. An introduction to these principles and their associated estimators is contained in Elementary Statistical Methods For Foresters, Agriculture Handbook #317 (literature cited).

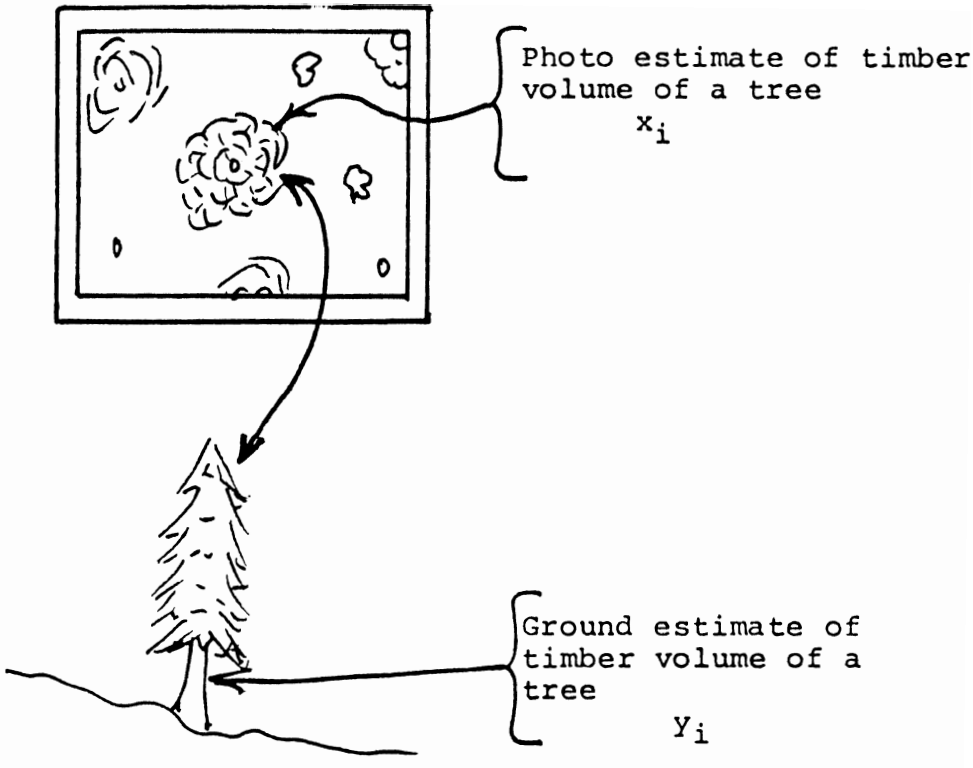

Figure 2; Multi-Phase Sampling

If conditions are such that a simple linear relationship exists between x and y, the ratio (λ) of the number of primary observations (n) to auxiliary observation (m) is estimated by the following. 3/

$$\lambda = \frac{n}{m} = \sqrt{\frac{1-\rho_{xy}^2}{\rho_{xy}^2} \frac{C_1}{C_2}}$$

ρ_{xy} = correlation between x and y.

C_1 = cost of obtaining and processing an observation of an auxiliary variable.

C_2 = cost of obtaining and processing on observation of the primary variable.

3/ L.C. Wensel, Wildland Resource Sampling (Draft), University of California, Berkeley, pages 4-75 to 4-79.

The total cost (C) of the collecting and processing of the inventory data is

$$C = C_1 m + C_2 n.$$

Therefore for a fixed budget

$$m = \frac{C}{C_1 + C_2 \lambda} \quad ; \quad n = m\lambda$$

The points to be made here are: 1)the allocation of effort in multi-phase sampling is a function of relative cost of the primary and auxiliary variable and 2)the strength of the relationship between the x and y observations (correlation for simple linear relationships). This is different than the multi-stage sampling procedure which depended on cost and variance rather than cost and correlation.

The above discussion of sample size has been restricted to fixed budget sample allocation. For procedures to optimize a given precision level, the reference by L.C. Wensel provides a good introductory discussion.

The two simple cases just describes are easily named. When multiple data levels and information extraction methods are used the sampling and estimation procedures become complex and the naming of the procedure becomes difficult. The sampling systems are combinations of multi-stage and multi-phase with multiple stratifications. Care must be taken to optimize the system and adequately describe the resultant sampling procedure, including the estimators of the characteristics of interest and the confidence bound for these estimates.

Literature Cited

B. Husch, C.I. Miller, and T. Beers; Forest Mensuration; 2nd Edition; The Ronald Press Company.

Freese, F.; Elementary Statistical Methods For Foresters; Agriculture Handbook #317; U.S. Department of Agriculture, Forest Service; Jan. 1967.

Wensel, L.C.; Wildland Resource Sampling (Draft); University of California, Berkeley; 1977.