

7.3 Stand Inventory using Double (or “Two Phase”) Sampling (Specifically, Double Sampling for a Mean-of-Ratios Estimate)

Consider using this technique in any situation where measuring X (e.g., basal area) is quite a bit cheaper relative to measuring Y (e.g., height or volume)

Steps in using this technique

- Conduct a survey in which a large sample of X is taken to approximate μ_x (phase 1, a.k.a. primary phase)
- Take a sample of the sample units where X was measured and measure Y also to provide the relationship between X and Y (phase 2, or secondary phase)
- Apply large sample mean of X to the relationship between X and Y to estimate the population mean of Y

Double sampling with mean-of-ratios estimation is useful when the relationship between Y and X is direct, linear, and through the origin

Some notation

n^1 denotes sample size for the large, or “primary,” first-phase sample

\bar{x}^1 denotes mean of auxiliary variable from the first-phase sample

n denotes sample size for the small, or “secondary,” second-phase sample

\bar{x} denotes mean of the auxiliary variable from the second-phase sample

\bar{y} denotes mean of the variable of interest from the second-phase sample

Then, the mean of y on a per-sample-unit basis is estimated by:

$$\bar{y}_{dsr} = \hat{R}\bar{x}^1$$

where:

dsr subscript refers to Double Sampling for Mean-of-Ratios Estimate

$$\hat{R} = \frac{1}{n} \sum_{i=1}^n \frac{y_i}{x_i}, \text{ and all other variables as defined previously for ratio estimation}$$

The estimate for the total of Y, is:

$$\hat{T}_{Ydsr} = N\bar{y}_{dsr}$$

How good is the estimate? A standard error is needed to answer this question.

There are two ways to do this, 1) when second phase is true subset of first phase, meaning the phases are dependent, or 2) when phases are independent

For dependent phases, the variance of the mean is:

$$S_{\bar{y}_{dsr}}^2 = \left(\frac{\bar{x}'}{\bar{x}} \right)^2 \left(\frac{S_y^2 + \hat{R}^2 S_x^2 - 2\hat{R}S_{xy}}{n} \right) \left(1 - \frac{n}{n'} \right) + \frac{S_y^2}{n'} \left(1 - \frac{n'}{N} \right)$$

where,

S_y^2 = variance of y in the second-phase sample,

S_x^2 = variance of x in the second-phase sample,

$$S_{xy} = \frac{\sum_{i=1}^n x_i y_i - \frac{1}{n} \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n-1} = \text{covariance of x \& y in sub-sample, and}$$

all other variables are defined as before

For independent phases, Bruce's method is used to compute Standard Error % of the mean of y:

$$SE\%_{\bar{y}} = \sqrt{SE\%_{\hat{R}}^2 + SE\%_{\bar{x}'}^2}$$

where,

$$SE\%_{\hat{R}}^2 = \left[\frac{SE_{\hat{R}}}{\hat{R}} (100) \right]^2$$

$$SE\%_{\bar{x}'}^2 = \left[\frac{SE_{\bar{x}'}}{\bar{x}'} (100) \right]^2$$

For variable-area plot sampling situations, we let

X = basal area, and

R = Volume to Basal Area Ratio, or VBAR

Example. The following data were collected during a stand inventory conducted to determine cubic-foot volume to a 4" top (CV4) per acre. A 20 BAF angle gauge was used for estimating structure from a large sample of trees, a 90 BAF was used for selecting VBAR trees. Average tariff # was calculated to equal 35.5 as determined from the VBAR sample trees.

Point 1		Point 2		Point 3	
DBH	VBAR	DBH	VBAR	DBH	VBAR
16	36.53	14	35.75	18	37.01
12		12	34.63	10	32.82
14		16		12	
18	37.01	10	32.82	14	
10		18		16	
		14		10	
		12		16	
TC= 5	36.77	TC= 7	34.40	TC= 7	34.92

$$\bar{X} (MTC) = 6.33 \quad S_X^2 = 1.33 \quad S_{\bar{X}'} = 0.66 \quad SE\%_{\bar{X}'} = \frac{0.66}{6.33}(100) = 10.53\%$$

$$\hat{R} (VBAR) = 35.31 \text{ ft}^3/\text{ft}^2 \quad S_{\hat{R}}^2 = 1.72 \quad S_{\hat{R}} = 0.76 \quad SE\%_{\hat{R}} = \frac{0.76}{35.31}(100) = 2.15\%$$

$$CV4/ac = MTC \times BAF \times VBAR = 4470.25 \text{ ft}^3/ac$$

$$SE\%_{\bar{y}} = \sqrt{SE\%_{\hat{R}}^2 + SE\%_{\bar{X}'}^2} = \sqrt{(10.53_{\hat{R}})^2 + (2.15_{\bar{X}'})^2} = 10.74\%$$

$$SE_{CV4/ac} = SE_{\bar{y}} = \frac{SE\%_{\bar{y}}}{100}(4470.25) = 480.10 \text{ ft}^3/ac$$

Note that both $n' = 3$ and $n = 3$ in this situation. This happens often when independent phases are used. Further, it might be argued that n is still "smaller" than n' in the sense that fewer trees than all possible the trees at a point are selected for measurement (i.e., a subset).

Confidence intervals can now be constructed in the usual way:

$$CV4/ac \pm t_{1-\alpha,df} \cdot SE_{CV4/ac}, \text{ or more generally, } \bar{y}_{dsr} \pm t_{1-\alpha,df} \cdot S_{\bar{y}_{dsr}}$$