Simple Random Sampling

n *items measured out of* N *possible items* (*sometimes* N *is infinite*)

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad \sum_{i=1}^{n} y_i \quad (\text{sum over all } n \text{ items})$$

$$\sum_{i=1}^{n} y_i^2 \; (\text{square each value and then add them})$$

$$s_y^2 = \frac{\sum y_i^2 - \left(\sum y_i\right)^2 / n}{n-1} \qquad s_{\bar{y}}^2 = \frac{s_y^2}{n}\left(\frac{N-n}{N}\right) \text{ without replacement;}$$

$$s_{\bar{y}}^2 = \frac{s_y^2}{n} \text{ with replacement or when } N \text{ is very large}$$

$$y_T = N \times \bar{y} \qquad s_{y_T}^2 = N^2 \times s_{\bar{y}}^2$$

(*when* $\bar{y}$ *is already on a per ha basis* (*e.g.*, m^3 *per ha, then use* A (*area in ha*) *instead of* N *in calculating* y_T *and* $s_{y_T}^2$)

$$CV = \frac{s_y}{\bar{y}} \times 100$$

95% *Confidence Intervals for the mean of the population*:

$$\bar{y} +/- \; t_{n-1,1-\alpha/2} \times s_{\bar{y}}$$

95% *Confidence Intervals for the total of the population*:

$$y_T +/- \; t_{n-1,1-\alpha/2} \times s_{y_T}$$

$$\text{Percent error achieved} = \frac{\text{CI width}}{\text{est. mean}} \times 100 = \frac{t_{n-1,1-\alpha/2} \times s_{\bar{y}}}{\bar{y}} \times 100 \text{ for a completed survey}$$

$$\text{Desired Percent error}(PE) = \frac{AE}{\bar{y}} \times 100 \text{ when } AE \text{ is given for a new survey}$$

Sample Size *without replacement*: *with replacement or very large* N:

$$n = \frac{1}{\frac{1}{N} + \frac{AE^2}{t^2 \times s_y^2}} = \frac{1}{\frac{1}{N} + \frac{PE^2}{t^2 \times CV^2}} \qquad n = \frac{t^2 \times s_y^2}{AE^2} = \frac{t^2 \times CV^2}{PE^2}$$

Systematic Sampling

Strip Sampling

- each strip is one observation, and the value for each strip is y_i in the simple random sampling formula, and n is the number of strips selected for measurement
- randomly select the first strip
- Intensity is calculated as:

$$I = \frac{n \times a}{A}$$

- Can calculate number of strips needed based on intensity desired, the size of the area (A), and the area under each plot (a)

Line Plot Sampling (or grid sampling)

- each plot is one observation, and the value for each plot is y_i in the simple random sampling formula, and n is the number of plots.

$$Sample\ Size:\ 1)\quad n = \frac{I \times A}{a}\quad or\quad 2)\ use\ simple\ random\ sampling\ equations\ for\ a\ specified\ AE\ or\ PE\ or\ 3)\ sample\ size\ is\ \text{limited by cost}$$

Spacing:

$$d(m) = \sqrt{\frac{A(m^2)}{n}}\qquad or\qquad L(m) = \frac{A(m^2)}{n \times B(m)}\quad or\quad B(m) = \frac{A(m^2)}{n \times L(m)}$$

Line Intersect Sampling

For each line with length L in metres, dij in cm, and θ ij in degrees:

$$y_i(m^3/ha) = \frac{\pi^2}{8 \times L}\sum_{j=1}^{mi} dij^2 \Big/ \cos(\theta\ ij)$$

$$y_i(stems/ha) == \frac{10000 \times \pi}{2 \times L}\sum_{j=1}^{mi} \frac{1}{lij \times \cos(\theta\ ij)}$$

then the y_i *are used in the simple random sampling equations to get an estimate of the mean, confidence intervals for this, and an estimate of the total with an associated confidence interval. The sample size, n, is the number of lines.*

Sampling for Proportions of Observations that have a particular characteristic

Single Observations, No stratification

$$\hat{p} = \frac{\text{number having the characteristic}}{\text{number of sampled}} = \frac{\sum_{i=1}^{n} y_i}{n}$$

$y_i = 1$ *if it has the characteristic and* 0 *otherwise*

N *is the total number of observations in the population*

Confidence intervals:

One way: normal approximation when np > 5 or nq > 5

$$\hat{q} = 1 - \hat{p} \qquad s_{\hat{p}} = \sqrt{\frac{\hat{p}\hat{q}}{n-1}\left(1 - \frac{n}{N}\right)} \qquad \hat{p} \pm z_{\frac{\alpha}{2}} \times s_{\hat{p}} + \frac{1}{2n}$$

For example, for a 95% confidence interval, z=1.96. For a 90% confidence interval, z=1.64. Another way: for 95 % confidence intervals, use the binomial graph.

Testing Hypotheses,

One proportion: Is the proportion more than (some value)? (Note: could also test if the proportion was less than a value)

$H_0: \; p = (\textit{some value})$

$H_0: \; p > (\textit{some value})$

$$z = \frac{\hat{p} - (\textit{some value})}{s_{\hat{p}}}$$

Compare to z critical from a normal distribution table. If z calculated is greater than z in the table, then reject H_0. *For example: z critical for* $\alpha = 0.05$ *is* 1.64

Comparing proportions from two populations:

$H_0: \; p_1 = p_2$

$H_0: \; p_1 > p_2$

$$s^2_{\hat{p}_1} = \frac{\hat{p}_1 \times \hat{q}_1}{n_1 - 1}\left(\frac{N_1 - n_1}{N_1}\right) \qquad s^2_{\hat{p}_2} = \frac{\hat{p}_2 \times \hat{q}_2}{n_2 - 1}\left(\frac{N_2 - n_2}{N_2}\right)$$

$$z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{s^2_{\hat{p}_1} + s^2_{\hat{p}2}}}$$

Compare to z critical from a normal distribution table. If z calculated is greater than z in the table, then reject H_0. *For example: z critical for* $\alpha = 0.05$ *is* 1.64

Sample size for estimating one proportion:

$$n = \frac{1}{\frac{1}{N} + \frac{AE^2}{z^2 \times \hat{p} \times \hat{q}}}$$

Proportions in clusters (e.g., plots, trucks, houses, or any group of items)

Two types of analysis:

1. Give equal weight to the proportion from each cluster. Calculate a proportion for each cluster, and treat each proportion as *yi* in simple random sampling equations

2. Give plots with more items more weight. Equations are then based on ratio of means sampling:

N=number of clusters in the population

n=number of clusters sampled

mi= number of items in cluster *i*

ai = number of items in cluster *i* that have the characteristic (e.g. dead, or whatever characteristic)

$$\overline{m} = \sum_{i=1}^{n} mi \Big/ n \quad \textit{the average number of items per cluster}$$

$$\overline{p} = \frac{\sum_{i=1}^{n} ai}{\sum_{i=1}^{n} mi} \quad \textit{the estimated proportion}$$

$$\textit{Need also}: \quad \sum_{i=1}^{n} ai^{2} \qquad \sum_{i=1}^{n} mi^{2} \qquad \sum_{i=1}^{n} ai\,mi = \sum_{i=1}^{n} ai \times mi$$

$$s_{\overline{p}}^{2} = \frac{(N-n)}{N} \times \frac{1}{n \times \overline{m}^{2}} \times \frac{\sum ai^{2} - 2\overline{p}\sum aimi + \overline{p}^{2}\sum mi^{2}}{n-1}$$

$$s_{\overline{p}} = \sqrt{s_{\overline{p}}^{2}}$$

$$95\,\% \quad \textit{Confidence Intervals for the proportion}:$$

$$\overline{p} +/- \quad t_{n-1,1-\alpha/2} \times s_{\overline{p}}$$

Stratified Random Sampling

For each stratum (use simple random sampling equations for data from the stratum):

n_h *items measured out of* N_h *possible items (sometimes* N_h *is infinite)*

$$\bar{y}_h = \frac{\sum_{i=1}^{n} y_{ih}}{n_h} \qquad \sum_{i=1}^{n} y_{ih} \ (sum\ over\ all\ n_h\ items) \qquad \sum_{i=1}^{n} y_{ih}^2\ (square\ each\ and\ sum)$$

$$s_{yh}^2 = \frac{\sum y_{ih}^2 - \left(\sum y_{ih}\right)^2 / n_h}{n_h - 1} \qquad s_{\bar{y}h}^2 = \frac{s_{yh}^2}{n_h}\left(\frac{N_h - n_h}{N_h}\right) \ or \ s_{\bar{y}h}^2 = \frac{s_{yh}^2}{n_h}$$

$$y_{Th} = N_h \times \bar{y}_h \qquad s_{y_{Th}}^2 = N_h^2 \times s_{\bar{y}h}^2 \qquad CV_h = \frac{s_{yh}}{\bar{y}_h} \times 100$$

(when $\bar{y}_h$ *is already on a per ha basis* $(e.g.\ m^3/ha)$*, then use* A_h *(area in ha) instead of* N_h *in calculating* y_{Th} *and* $s_{y_{Th}}^2$*)*

95% *Confidence Intervals for the mean for Stratum h*:

$$\bar{y}_h +/- \ t_{n-1,1-\alpha/2} \times s_{\bar{y}h}$$

95% *Confidence Intervals for the total for Stratum h*:

$$y_{Th} +/- \ t_{n-1,1-\alpha/2} \times s_{y_{Th}}$$

Overall (putting strata together to get estimates for the entire population):

$W_h = \frac{N_h}{N}$ *or* $W_h = \frac{A_h}{A}$ *where* W_h *is the relative size of Stratum h*

$$\bar{y}_{ST} = \sum_{h=1}^{L} W_h \times \bar{y}_h = \sum_{h=1}^{L} \frac{N_h}{N} \times \bar{y}_h \qquad y_{TST} = N \times \bar{y}_{ST}$$

$$s_{\bar{y}_{ST}}^2 = \sum_{h=1}^{L} W_h^2 \times s_{\bar{y}_h}^2 = \sum_{h=1}^{L} \frac{N_h^2}{N^2} \times s_{\bar{y}_h}^2 \qquad s_{y_{TST}}^2 = N^2 \times s_{\bar{y}_{ST}}^2$$

(if $\bar{y}_{st}$ *is expressed on a per ha basis* $(e.g.\ m^3/ha)$ *then replace N by A (area for population) to obtain* y_{TST} *and* $s_{y_{TST}}^2$

95% Confidence Intervals for the mean for the population:

$$\bar{y}_{ST} +/- \ t_{EDF} \times s_{\bar{y}_{ST}}$$

95% Confidence Intervals for the total for the population:

$$y_{TST} +/- \ t_{EDF} \times s_{y_{TST}}$$

where EDF is somewhere between the smallest $n_h - 1$ *up to the sum of the* $n_h - 1$ *for all strata combined. EDF can be calculated by*:

$$EDF = \frac{\left(s_{\bar{y}_{ST}}^2\right)^2}{\sum_{h=1}^{L} \frac{\left(W_h^2 \times s_{\bar{y}_h}^2\right)^2}{n_h - 1}} = \frac{\left(s_{\bar{y}_{ST}}^2\right)^2}{\sum_{h=1}^{L} \frac{\left(\frac{N_h^2}{N^2} \times s_{\bar{y}_h}^2\right)^2}{n_h - 1}}$$

$$Percent\ error\ achieved = \frac{CI\ width}{est.\ mean} \times 100 = \frac{t_{EDF} \times s_{\bar{y}_{ST}}}{\bar{y}_{ST}} \times 100 \ for\ a\ completed\ survey$$

Stratified Random Sampling (Continued)

Sample Size Calculations:

$$Equal \quad allocation:$$
$$n_h = n / L$$

$$where \; n \; is \; calculated \; as: \quad without \; replacement$$

$$n = \frac{L \times t^2 \times \sum_{h=1}^{L} W^2{}_h \times s^2{}_h}{AE^2 + t^2 \times \frac{\sum_{h=1}^{L} W_h \times s^2{}_h}{N}} = \frac{L \times t^2 \times \sum_{h=1}^{L} N^2{}_h \times s^2{}_h}{N^2 \times AE^2 + t^2 \times \sum_{h=1}^{L} N_h \times s^2{}_h}$$

$$for \; with \; replacement \; or \; when \; N \; is \; very \; large$$

$$n = \frac{L \times t^2 \times \sum_{h=1}^{L} W^2{}_h \times s^2{}_h}{AE^2} = \frac{L \times t^2 \times \sum_{h=1}^{L} N^2{}_h \times s^2{}_h}{N^2 \times AE^2}$$

$$\Pr oportional \quad allocation:$$
$$n_n = \frac{N_h}{N} \times n = W_h$$

$$where \; n \; is \; calculated \; as: \quad without \; replacement$$

$$n = \frac{t^2 \times \sum_{h=1}^{L} W_h \times s^2{}_h}{AE^2 + t^2 \times \frac{\sum_{h=1}^{L} W_h \times s^2{}_h}{N}} = \frac{N \times t^2 \times \sum_{h=1}^{L} N_h \times s^2{}_h}{N^2 \times AE^2 + t^2 \times \sum_{h=1}^{L} N_h \times s^2{}_h}$$

$$for \; with \; replacement \; or \; when \; N \; is \; very \; large$$

$$n = \frac{t^2 \times \sum_{h=1}^{L} W_h \times s^2{}_h}{AE^2} = \frac{N \times t^2 \times \sum_{h=1}^{L} N_h \times s^2{}_h}{N^2 \times AE^2}$$

Stratified Random Sampling (Continued)

Neyman allocation :

$$n_h = n \times \frac{W_h \times s_h}{\sum_{h=1}^{L} W_h \times s_h}$$

where n is calculated as : without replacement

$$n = \frac{t^2 \times \left(\sum_{h=1}^{L} W_h \times s_h \right)^2}{AE^2 + t^2 \times \frac{\sum_{h=1}^{L} W_h \times s^2{}_h}{N}} = \frac{t^2 \times \left(\sum_{h=1}^{L} N_h \times s_h \right)^2}{N^2 \times AE^2 + t^2 \times \sum_{h=1}^{L} N_h \times s^2{}_h}$$

for with replacement or when N is very large

$$n = \frac{t^2 \times \left(\sum_{h=1}^{L} W_h \times s_h \right)^2}{AE^2} = \frac{t^2 \times \left(\sum_{h=1}^{L} N_h \times s_h \right)^2}{N^2 \times AE^2}$$

Optimum allocation :

$$n_h = n \times \frac{W_h \times s_h / \sqrt{c_h}}{\sum_{h=1}^{L} W_h \times s_h / \sqrt{c_h}}$$

where n is calculated as : without replacement

$$n = \frac{t^2 \times \left(\sum_{h=1}^{L} W_h \times s_h \sqrt{c_h} \right) \left(\sum_{h=1}^{L} \frac{W_h \times s_h}{\sqrt{c_h}} \right)}{AE^2 + t^2 \times \frac{\sum_{h=1}^{L} W_h \times s^2{}_h}{N}} = \frac{t^2 \times \left(\sum_{h=1}^{L} N_h \times s_h \sqrt{c_h} \right) \left(\sum_{h=1}^{L} \frac{N_h \times s_h}{\sqrt{c_h}} \right)}{N^2 \times AE^2 + t^2 \times \sum_{h=1}^{L} N_h \times s^2{}_h}$$

for with replacement or when N is very large :

$$n = \frac{t^2 \times \left(\sum_{h=1}^{L} W_h \times s_h \sqrt{c_h} \right) \left(\sum_{h=1}^{L} \frac{W_h \times s_h}{\sqrt{c_h}} \right)}{AE^2} = \frac{t^2 \times \left(\sum_{h=1}^{L} N_h \times s_h \sqrt{c_h} \right) \left(\sum_{h=1}^{L} \frac{N_h \times s_h}{\sqrt{c_h}} \right)}{N^2 \times AE^2}$$

Ratio/Regression Sampling

$n = number\ of\ items\ with\ both\ x\ and\ y\ measures;$
$N = number\ of\ items\ in\ the\ population \quad x\ is\ the\ auxilliary\ variable$
$\mu_x\ and\ \tau_x\ known,\ as\ x\ is\ measured\ on\ all\ N\ items$

Using Mean of Ratios

$$\sum_{i=1}^{n} y_i / x_i = \sum_{i=1}^{h} r_i \qquad \sum_{i=1}^{n} r_i^2 \qquad \bar{r} = \frac{\sum_{i=1}^{h} r_i}{n}$$

$$s_r^2 = \frac{\sum r_i^2 - \left(\sum r_i\right)^2 / n}{n-1} \qquad s_r = \sqrt{s_r^2} \qquad CV = \frac{s_r}{\bar{r}} \times 100$$

$$s_{\bar{r}}^2 = \frac{s_r^2}{n} \times \left(\frac{N-n}{N}\right) \qquad s_{\bar{r}} = \sqrt{s_{\bar{r}}^2}$$

$$\bar{y}_r = \bar{r} \times \mu_x \qquad y_{Tr} = N \times \bar{y}_r$$

$$s_{\bar{y}_r} = s_{\bar{r}} \times \mu_x \qquad s_{y_{Tr}} = N \times s_{\bar{y}_r}$$

$(when\ \bar{y}_r\ is\ already\ on\ a\ per\ ha\ basis\ (e.g.\ m^3/ha),\ then\ use\ A(area\ in\ ha)$
$instead\ of\ N\ in\ calculating\ y_{Tr}\ \ nd\ s_{y_{Tr}})$

$95\%\ Confidence\ Intervals\ for\ the\ mean\ of\ the\ population:$

$$\bar{y}_r +/- \ t_{n-1,1-\alpha/2} \times s_{\bar{y}_r}$$

$95\%\ Confidence\ Intervals\ for\ the\ total\ of\ the\ population:$

$$y_{Tr} +/- \ t_{n-1,1-\alpha/2} \times s_{y_{Tr}}$$

$$Achieved\ Percent\ error = \frac{CI\ width}{est.mean} \times 100 = \frac{t_{n-1,1-\alpha/2} \times s_{\bar{y}_r}}{\bar{y}_r} \times 100$$

$Sample\ Size\ without\ replacement:$

$$n = \frac{1}{\frac{1}{N} + \frac{PE^2}{t^2 \times CV^2}}$$

$with\ replacement\ or\ very\ large\ N:$

$$n = \frac{t^2 \times CV^2}{PE^2}$$

Ratio/Regression Sampling (continued)

Using Ratio of Means

$$R = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = \frac{\bar{y}}{\bar{x}} \qquad s_R^{\,2} = \left(\frac{1}{n \times \mu_x^{\,2}}\right) \times \left(\frac{\sum y_i^{\,2} - 2 \times R \times \sum x_i y + R^2 \sum x_i^{\,2}}{n-1}\right) \times \left(\frac{N-n}{N}\right)$$

$$s_R = \sqrt{s_R^{\,2}}$$

$$\bar{y}_R = R \times \mu_x \qquad y_{TR} = N \times \bar{y}_R$$

$$CV = \frac{s_1}{\bar{y}_R} \times 100 = \frac{\sqrt{\sum y_i^{\,2} - 2 \times R \times \sum x_i y + R^2 \sum x_i^{\,2} \big/ (n-1)}}{\bar{y}_R} \times 100$$

$$s_{\bar{y}_R} = s_R \times \mu_x \qquad s_{y_{TR}} = N \times s_{\bar{y}_R}$$

(*when* $\bar{y}_R$ *is already on a per ha basis* (*e.g.* m^3/ha), *then use* A(*area in ha*) *instead of* N *in calculating* y_{TR} *nd* $s_{y_{TR}}$)

95% *Confidence Intervals for the mean of the population*:

$$\bar{y}_R \;+/-\; t_{n-1,1-\alpha/2} \times s_{\bar{y}_R}$$

95% *Confidence Intervals for the total of the population*:

$$y_{TR} \;+/-\; t_{n-1,1-\alpha/2} \times s_{y_{TR}}$$

$$Percent \quad error = \frac{CI\ width}{est.mean} \times 100 = \frac{t_{n-1,1-\alpha/2} \times s_{\bar{y}_R}}{\bar{y}_R} \times 100$$

Sample Size *without replacement*:

$$n = \frac{1}{\dfrac{1}{N} + \dfrac{AE^2}{t^2 \times s_1^{\,2}}} = \frac{1}{\dfrac{1}{N} + \dfrac{PE^2}{t^2 \times CV^2}}$$

with replacement or very large N:

$$n = \frac{t^2 \times s_1^{\,2}}{AE^2} = \frac{t^2 \times CV^2}{PE^2}$$

Ratio/Regression Sampling (continued)

Using Regression

$$Need: \quad \sum_{i=1}^{n} y_i \qquad \sum_{i=1}^{n} y_i^{\,2} \qquad \bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$$

$$\sum_{i=1}^{n} x_i \qquad \sum_{i=1}^{n} x_i^{\,2} \qquad \sum_{i=1}^{n} x_i y_i \qquad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$SPXY = \sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right)/n \qquad SSX = \sum x_i^{\,2} - \left(\sum x_i\right)^2/n$$

$$SSY = \sum y_i^{\,2} - \left(\sum y_i\right)^2/n$$

then:

$$slope: \quad b_1 = \frac{SPXY}{SSX} \qquad \text{int}\,ercept: \quad b_0 = \bar{y} - b_1 \times \bar{x}$$

$$\bar{y}_{lr} = b_0 + b_1 \times \mu_x \qquad y_{T\,lr} = N \times \bar{y}_{lr}$$

$$s_{\bar{y}_{lr}} = SE_E \times \sqrt{\frac{1}{n} + \frac{(\mu_x - \bar{x})^2}{SSX}} \times \sqrt{\left(\frac{N-n}{N}\right)} \qquad SE_E = \sqrt{\frac{SS_{RES}}{n-2}} \qquad SS_{RES} = SSY - b_1 \times SPXY$$

$$s_{y_{T\,lr}} = N \times s_{\bar{y}_{lr}}$$

(*when* $\bar{y}_{lr}$ *is already on a per ha basis* (*e.g.* m^3/ha), *then use* A(*area in ha*) *instead of* N *in calculating* $y_{T\,lr}$ *and* $s_{y_{T\,lr}}$)

95% Confidence Intervals for the mean of the population:

$$\bar{y}_{lr} \ +/- \ t_{n-1,1-\alpha/2} \times s_{\bar{y}_{lr}}$$

95% Confidence Intervals for the total of the population:

$$y_{T\,lr} \ +/- \ t_{n-1,1-\alpha/2} \times s_{y_{T\,lr}}$$

$$Achieved \ Percent \ error = \frac{CI\ width}{est.mean} \times 100 = \frac{t_{n-1,1-\alpha/2} \times s_{\bar{y}_{lr}}}{\bar{y}_{lr}} \times 100$$

Sample Size: (*not* covered *in FRST* 339)

Double Sampling

Phase 1: Large sample for x

$$n' \text{ observations on } x \qquad \bar{x}' = \sum_{i=1}^{n'} x_i \Big/ n' \qquad s^2{}_x{}' = \frac{\sum_{i=1}^{n'} x_i{}^2 - \left(\sum_{i=1}^{n'} x_i\right)^2 / n'}{n'-1}$$

Phase 2: Smaller sample, measure y also.

n observations on x and on y

$$\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n} \qquad s^2{}_y = \frac{\sum y_i{}^2 - \left(\sum y_i\right)^2 / n}{n-1} \qquad \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$s^2{}_x = \frac{\sum x_i{}^2 - \left(\sum x_i\right)^2 / n}{n-1} \qquad s_{xy} = \frac{\sum x_i y_i - \left(\sum x_i\right)\left(\sum y_i\right) / n}{n-1}$$

Double Sampling: Using ratio of means (relationship between y and x goes through zero):

$$\hat{R}_{2-P} = \frac{\sum_{i=1}^{n} y_i}{\sum_{i=1}^{n} x_i} = \frac{\bar{y}}{\bar{x}} \qquad \bar{y}_{2-P} = \hat{R}_{2-P} \times \bar{x}' \qquad y_{T\,2-P} = N \times \bar{y}_{2-P}$$

$$s^2{}_{\bar{y}_{2-P}} = \frac{s^2{}_y + \hat{R}_{2-P}{}^2 \times s^2{}_x - 2 \times \hat{R}_{2-P} \times s_{xy}}{n} + \frac{2 \times \hat{R}_{2-P} \times s_{xy} - \hat{R}_{2-P}{}^2 \times s^2{}_x{}'}{n'} - \frac{s^2{}_y}{N}$$

NOTE: last term drops out if N is very large or if sampling is with replacement.

Can use $s^2{}_x$ (from Phase 2) instead of $s^2{}_x{}'$ (from Phase 1)

$$s^2{}_{y_{T\,2-P}} = N^2 \times s^2{}_{\bar{y}_{2-P}}$$

(*when $\bar{y}_{2-P}$ is already on a per ha basis (e.g. m^3/ha), then use A (area in ha) instead of N in calculating $y_{T\,2-P}$ and $s_{y_{T\,2-P}}$*)

95% *Confidence Intervals for the mean for the population:*

$$\bar{y}_{2-P} \; +/- \; t_{n-1,1-\alpha/2} \times s_{\bar{y}_{2-P}}$$

95% *Confidence Intervals for the total for the population:*

$$y_{T\,2-P} \; +/- \; t_{n-1,1-\alpha/2} \times s_{y_{T\,2-P}}$$

$$\text{Percent error achieved} = \frac{CI\ width}{est.\ mean} \times 100 = \frac{t_{n-1,1-\alpha/2} \times s_{\bar{y}_{2-P}}}{\bar{y}_{2-P}} \times 100$$

Double Sampling

Double Sampling: Using regression (relationship between y and x can go through something other than zero)

$$b = \frac{s_{xy}}{s_x^{\ 2}} \quad \textit{where } b \textit{ is the estimated slope between } y \textit{ and } x$$

$$\bar{y}_{lr\,2-P} = \bar{y} - b(\bar{x} - \bar{x}') \qquad y_{T\ lr\ 2-P} = N \times \bar{y}_{lr\,2-P}$$

$$s^2_{\bar{y}_{lr2-P}} = \left(\frac{n-1}{n-2}\right) \times \left(\frac{s^2{}_y - (s_{xy})^2 / s_x^{\ 2}}{n}\right) \times \left(1 + \frac{1}{n} - \frac{1}{n'}\right) \ + \left(b^2 \times \frac{s_x^{\ 2}}{n'}\right)$$

$$s^2{}_{y_{T\,lr\,2-P}} = N^2 \times s^2{}_{\bar{y}_{lr2-P}}$$

(*when* $\bar{y}_{lr2-P}$ *is already on a per ha basis* $(e.g.\ m^3 / ha)$, *then use* A (*area in ha*) *instead of* N *in calculating* $\bar{y}_{lr\,2-P}$ *and* $s^2{}_{y_{T\,lr\,2-P}}$)

95% *Confidence Intervals for the mean of the population*:

$$\bar{y}_{lr\,2-P} \ +/- \ \ t_{n-1,1-\alpha/2} \times s_{\bar{y}_{lr2-P}}$$

95% *Confidence Intervals for the total of the population*:

$$y_{T\ lr\ 2-P} \ +/- \ \ t_{n-1,1-\alpha/2} \times s_{y_{T\,lr\,2-P}}$$

$$\textit{Achieved} \quad \textit{Percent error} = \frac{\textit{CI width}}{\textit{est.mean}} \times 100 = \frac{t_{n-1,1-\alpha/2} \times s_{\bar{y}_{lr2-P}}}{\bar{y}_{lr\,2-P}} \times 100$$