i-Tree Sample Total and Variance Estimates

i-Tree Eco uses the following formulas to estimate variance based on plot samples. These formulas were provided by US Forest Service statistician David Randall. Stratification is often by land use, but could be any stratum as defined by the user. Standard error for each estimator is the square root of its variance.

Notation:

A_r = total area in stratum r

A. = total area in all strata = $\sum A_r$

a_r = total area of samples in stratum r

$\bar{a}_r$ = mean area of samples in stratum r = $\sum_{i=1}^{n_r} {a_{ri}}/{n_r}$ where n_r = number of samples

Sample Variance

$$s^2(y)_r = \sum_{i=1}^{n_r} {(y_{r_i} - \overline{y}_r)^2}/{(n_r - 1)}$$ = sample variance of variable "y" in stratum "r"

$$s(xy)_r = \sum_{i=1}^{n_r} (y_{r_i} - \overline{y}_r)(x_{r_i} - \overline{x}_r)/(n_r - 1)$$

= sample covariance of variable "y" and "x" in stratum "r"

Where: n_{r_i} = number of trees within sample (plot) "i" of stratum "r"

$y_{r_{i_j}}$ = value of variable "y" (for which stratum mean is being estimated) in subsample "j" of sample (plot) "i" of stratum "r"

$$y_{r_i} = \sum_{j=1}^{n_{r_i}} y_{r_{i_j}}$$ = plot total

$$\hat{\overline{y}}_r = \sum_{i=1}^{n_r} y_{r_i}/n_r$$

Estimators for each stratum (e.g., land use type):

a) Stratum sample means (it is assumed that the variable is expressed as an absolute measurement [e.g. a count, an area, a mass, etc.] and not as a relative measurement [e.g. a proportion, %, etc.]:

$$\text{estimator} = \hat{\bar{y}}_r = \sum_{i=1}^{n_r} y_{r_i}/n_r$$

$$\text{variance} = \widehat{Var}\left(\hat{\bar{y}}_r\right) = \frac{(1-f_r)}{n_r} s^2(y)_r$$

where: n_r = number of samples in stratum "r"
f_r = sampling fraction in stratum "r" = ${a_r}/{A_r}$

b) Stratum mean expressed on per/area basis (stratum total/stratum area) (e.g., “y” in number of stems and “a” is acres, yields stems/acre):

$$\text{estimator} = \left(\frac{1}{\bar{a}_r}\right)\hat{\bar{y}}_r = (\hat{M}_{area})_r$$

$$\text{variance} = \left(\frac{1}{\bar{a}_r}\right)^2 \widehat{Var}\left(\hat{\bar{y}}_r\right) = \widehat{Var}\left(\hat{M}_{area}\right)_r$$

where:
a_r = total area of samples in stratum “r”
a_r = mean area of samples in stratum “r” = $\sum_{i=1}^{n_r} a_{r_i}/n_r$

c) Stratum total per stratum area expressed as a percent (stratum total [in area]/stratum area); “y” and “a” must be in identical units of area; e.g. % cover:

$$\text{estimator} = \left(\frac{100}{\bar{a}_r}\right)\hat{\bar{y}}_r = (\hat{M}_{\%})_r$$

$$\text{variance} = \left(\frac{100}{\bar{a}_r}\right)^2 \widehat{Var}\left(\hat{\bar{y}}_r\right) = \widehat{Var}\left(\hat{M}_{\%}\right)_r$$

d) Total over entire stratum, where total has meaning (“A” and “a” must be identical units of area):

$$\text{estimator} = \left(\frac{A_r}{\bar{a}_r}\right)\hat{\bar{y}}_r = (\hat{T})_r$$

$$\text{variance} = \left(\frac{A_r}{\bar{a}_r}\right)^2 \widehat{Var}\left(\hat{\bar{y}}_r\right) = \widehat{Var}\left(\hat{T}\right)_r$$

where:
A_r = total area in stratum “r”

e) Ratio of Totals of Two Random Variables (stratum total of “y”/stratum total of “x”); only those plots in which both variables were measured should be used in the estimation (e.g., “y”= stems of maple, x = all stems, to yield proportion of stems that are maple, or “x” = stems of oak, to yield ratio of maple to oak):

$$\text{estimator} = \frac{\sum_{i=1}^{n_r} y_{r_i}}{\sum_{i=1}^{n_r} x_{r_i}} = \frac{\hat{\bar{y}}_r}{\hat{\bar{x}}_r} = \frac{(\hat{T}_y)_r}{(\hat{T}_x)_r} = (\hat{R})_r$$

variance = $\frac{(1-f_r)}{n_r}\left(\frac{1}{\hat{\bar{\bar{x}}}_r}\right)\left[s^2(y)_r + (\hat{R})_r^2 s^2(x)_r - 2(\hat{R})_r s(xy)_r\right] = \widehat{Var}(\hat{R})_r$

f) Subsample mean over entire stratum (where subsample here refers to one or more measurements made <u>within</u> the same sample unit, i.e., within the same <u>plot</u>); an example would be tree condition, where this is measured separately for each tree in the sample plot:

estimator = $\frac{\sum_{i=1}^{n_r} y_{r_i}}{\sum_{i=1}^{n_r} n_{r_i}} = (\hat{M})_r$

variance = $\frac{(1-f_r)}{n_r}\left(\frac{n_r}{\sum_{i=1}^{n_r} n_{r_i}}\right)^2 \left[s^2(y_{r_i}) + (\hat{M})_r^2 s^2(n_{r_i}) - 2(\hat{M})_r s(y_{r_i} n_{r_i})\right] = \widehat{Var}(\hat{M})_r$

Note for the above estimator and variance equations: n_r is included only on plots for which y_{r_i} is defined (in the case of tree condition, trees must be present). On the other hand, a_r and f_r are calculated for all observed samples (plots), regardless of whether y_{r_i} is defined.

<u>Population estimators that combine all strata (e.g., city total)</u>:

a) Sample mean over all strata (a mean of stratum means, weighted by stratum area):

estimator = $\hat{\bar{\bar{y}}}_{\bullet} = \left[\sum_{r=1}^{S} A_r\, (\hat{\bar{\bar{y}}}_r)\right] / A$

variance = $\widehat{Var}(\hat{\bar{\bar{y}}}_{\bullet}) = \left[\sum_{r=1}^{S} (A_r)^2 \widehat{Var}(\hat{\bar{\bar{y}}}_r)\right] / A^2$

where:

$A.$ = total area in all strata $= \sum_{r=1}^{S} A_r$

b) All strata total expressed on per/area basis (all-stratum total/all-stratum area):

estimator = $\left[\sum_{r=1}^{S} A_r\, (\hat{M}_{area})_r\right] / A = (\hat{M}_{area})_{\bullet}$

variance = $\left[\sum_{r=1}^{S} \left(\frac{A_r}{\bar{\bar{a}}_r}\right)^2 \widehat{Var}(\hat{\bar{\bar{y}}}_r)\right] / A^2 = \left[\sum_{r=1}^{S} (A_r)^2 \widehat{Var}(\hat{M}_{area})_r\right] / A^2 = \widehat{Var}(\hat{M}_{area})_{\bullet}$

c) All strata total expressed as percent (all-stratum total [in units of area]/all-stratum area); “y” and “a” must be in identical units of area:

estimator = $\left[\sum_{r=1}^{S} A_r\, (\hat{M}_{\%})_r\right] / A = (\hat{M}_{\%})_{\bullet}$

variance = $\left[\sum_{r=1}^{S}\left(\frac{100A_r}{\overline{a}_r}\right)^2 \widehat{Var}(\hat{\overline{y}}_r)\right] \Big/ A^2 = \left[\sum_{r=1}^{S}(A_r)^2\widehat{Var}(\hat{M}_{\%})_r\right] \Big/ A^2 = \widehat{Var}\left(\hat{M}_{\%}\right)_\bullet$

d) Total over all strata ("A" and "a" must be in identical units of area):

estimator = $\sum_{r=1}^{S}(\hat{T})_r = (\hat{T})_\bullet$

variance = $\sum_{r=1}^{S}\left(\frac{A_r}{\overline{a}_r}\right)^2 \left[\widehat{Var}(\hat{\overline{y}}_r)\right] = \sum_{r=1}^{S}\widehat{Var}(\hat{T})_r = (A^2)\widehat{Var}\left(\hat{M}_{area}\right)_\bullet = \widehat{Var}\left(\hat{T}\right)_\bullet$

e) Ratio of the totals of two random variables (all-stratum total of "y"/all-stratum total of "x"):

estimator = $\frac{\sum_{r=1}^{S}\left(\frac{A_r}{\overline{a}_r}\right)\hat{\overline{y}}_r}{\sum_{r=1}^{S}\left(\frac{A_r}{\overline{a}_r}\right)\hat{\overline{x}}_r} = \frac{\sum_{r=1}^{S}(\hat{T}_y)_r}{\sum_{r=1}^{S}(\hat{T}_x)_r} = (\hat{R})_\bullet$

variance = $\left[\sum_{r=1}^{S}\left(\frac{A_r}{\overline{a}_r}\hat{\overline{x}}_r\right)^2 \left[\widehat{Var}\left(\hat{R}\right)_r\right]\right] \Big/ \left[\sum_{r=1}^{S}\left(\frac{A_r}{\overline{a}_r}\hat{\overline{x}}_r\right)\right]^2 = \widehat{Var}(\hat{R})_\bullet$

f) Subsample mean over all strata, i.e. estimated ratio of total of all possible $y_{r_{i_j}}$ (over all strata) to total the count of such values:

estimator = $\frac{\sum_{r=1}^{S}\left[\left(\frac{A_r}{a_r}\right)\sum_{i=1}^{n_r} y_{r_i}\right]}{\sum_{r=1}^{S}\left[\left(\frac{A_r}{a_r}\right)\sum_{i=1}^{n_r} n_{r_i}\right]} = (\hat{M})_\bullet$

variance = $\left[\sum_{r=1}^{S}\left(\frac{A_r}{a_r}\sum_{i=1}^{n_r} n_{r_i}\right)^2 \left[\widehat{Var}\left(\hat{M}\right)_r\right]\right] \Big/ \left[\sum_{r=1}^{S}\left(\frac{A_r}{a_r}\sum_{r=1}^{S} n_{r_i}\right)\right]^2 = \widehat{Var}(\hat{M})_\bullet$