[redacted]

STT 351
Section 1

HW: 4, 6, 8, 12, 14, 16, 18

4. The published reference would be considered a benchmark because the values were already known and the programmer used the previously published data to compare with his/her own data.

6. I would use cluster sampling because the population doesn't have to be known. I would take a cluster of the trees and then use the SRS method to select a random sample.

8. a Yes both methods are capable of generating a random sample. Researcher A is choosing trees at random and then decreasing the sample size at random as well. Rescar B is using the integer method without replacement to choose the trees.

b This is called stratified sampling. The different sides of the hill are split up into different strata and then the SRS method is used to select a random sample

12. a) 10 equal strata
1000 smaller rectangles
$N = 10 \cdot 1000 = 10{,}000$

b) $n = \dfrac{1}{(.03/1.65)^2 + 1/1000}$

$n = 2322.46$

$n_1 = n_2 = n_3 = n_4 = \ldots = n_{10} = 232$

$n = \dfrac{1}{\left(\frac{.03}{1.65\sigma}\right)^2 + \frac{1}{N}}$ you needed to take $\sigma = 1$ since it's not specified.

c) $p_1 = \dfrac{5}{232} = .022$ $p_2 = 4/232 = .017$ $p_3 = 7/232 = .03$

$p_4 = 6/232 = .026$ $p_5 = 3/232 = .013$ $p_6 = 9/232 = .039$

$p_7 = 5/232 = .022$ $p_8 = 6/232 = .026$ $p_9 = 2/232 = .009$

$p_{10} = 8/232 = .034$

$p_{str} = .022\left(\dfrac{1000}{10000}\right) + .017\left(\dfrac{1}{10}\right) + .03\left(\dfrac{1}{10}\right) + .026\left(\dfrac{1}{10}\right) + .013\left(\dfrac{1}{10}\right)$

$+ .039\left(\dfrac{1}{10}\right) + .022\left(\dfrac{1}{10}\right) + .026\left(\dfrac{1}{10}\right) + .009\left(\dfrac{1}{10}\right) + .034\left(\dfrac{1}{10}\right)$

$p_{str} = .0238$

Standard error

$$\sqrt{\frac{1}{10{,}000^2}\left(1000^2\right)\left(\frac{1000-232}{1000}\right)}$$

$= .0876$

$$= .0876\left[\sqrt{\frac{.0022(1-.0022)}{231}}+\sqrt{\frac{.017(1-.017)}{231}}+\sqrt{\frac{.03(1-.03)}{231}}+\sqrt{\frac{.026(1-.026)}{231}}+\sqrt{\frac{.013(1-.013)}{231}}+\sqrt{\frac{.039(1-.039)}{231}}+\sqrt{\frac{.022(1-.022)}{231}}+\sqrt{\frac{.026(1-.026)}{231}}+\sqrt{\frac{.009(1-.009)}{231}}+\sqrt{\frac{.034(1-.034)}{231}}\right]$$

$= .0876[.0031 + .0085 + .011 + .0105 + .0075 + .0127 + .0097 + .0105 + .0062 + .0119]$

$= .008$

14. $w_i = \dfrac{N_i\sigma_i/\cancel{\sqrt{c_i}}}{\frac{N_1\sigma_1}{\cancel{\sqrt{c_1}}} + \frac{N_k\sigma_k}{\cancel{\sqrt{c_k}}}}$ cancel costs so

$n_i = n\left(\dfrac{N_i\sigma_i}{\sum_{i=1}^{k} N_i\sigma_i}\right)$

b) $w_i = \dfrac{N_i\sigma_i/\cancel{\sqrt{c_i}}}{\frac{N_i\sigma_i}{\cancel{\sqrt{c_i}}} + N_k\sigma_k/\cancel{\sqrt{c_k}}}$ cancel costs & σ so

$n_i = n\left(\dfrac{N_i}{N}\right)$

16. If you wanted the confidence interval to be 99% you would use a value between 2.57 & 2.58. Using a higher confidence level would give you a greater degree of reliability.

18. Replicated measurements help to ~~eliminate~~ reduce ~~biases~~ because several measurements are averaged. Also the variations between the repeated measurements gives an estimate of experimental error.

sampling question