FO-2213 Forest Measurements

Topic 15: Confidence Intervals and Sample Size Estimation

Statistical Formulae:

Variance: (working) $$s^2 = \frac{\left[\sum_{i=1}^{n} x_i^2 - \frac{(\sum_{i=1}^{n} x_i)^2}{n}\right]}{(n - 1)}$$ Or $$s^2 = \frac{\left[\sum_{i=1}^{n} x_i^2 - \bar{x}\sum_{i=1}^{n} x_i\right]}{(n - 1)}$$

Standard Error of the Mean: $$s_{\bar{x}} = \sqrt{\frac{s^2}{n}\left(1 - \frac{n}{N}\right)}$$

(1- ∝) Confidence Interval: $$\bar{x} \pm (t_{\alpha,\, n-1})s_{\bar{x}}$$

Sampling Error: $$SE\% = \left(\frac{t_{\alpha,\, n-1}\ s_{\bar{x}}}{\bar{x}}\right)100\%$$

Interpretation of the (1- α) Confidence Interval: (if α=0.05; then (1- α**)** = .95 or 95%)
If the sampling procedure is repeated an infinite number of times, I am **(1-α)**% confident that the true mean (μ) is within the interval $\bar{x} - (t_{\alpha,\, n-1})s_{\bar{x}}$ and $\bar{x} + (t_{\alpha,\, n-1})s_{\bar{x}}$.

The confidence interval does NOT mean the sample mean is 95% accurate or precise. It merely implies that unless a 1 in 20 chance has occurred, the confidence interval incorporates the true mean value of the sampled population.

The confidence interval and accompanying probability statements account for <u>sampling variation only</u>. It is assumed that:

1. sampling procedures are unbiased,
2. field measurements are without error, and
3. no computational mistakes were made.

<u>Sampling procedures are unbiased</u> means that n samples were randomly selected from a normally distributed population.

<u>Field measurements are taken/recorded without error</u> means there was no instrument error, the measurement was taken properly and appropriately, and the data were recorded without error.

<u>No computational mistakes were made</u> means appropriate statistical computations were performed without mathematical error.

The t-value is obtained from a table at the α probability level and (n-1) degrees of freedom. One degree of freedom is lost because the sample mean $\bar{x}$ was computed and is known.

The Distribution of t (abbreviated table)

	Probability = α								
df	**0.5**	**0.4**	**0.3**	**0.2**	**0.1**	**0.05**	**0.02**	**0.01**	**0.001**
1	1.000	1.376	1.963	3.078	6.314	12.706	31.821	63.657	636.619
2	0.819	1.061	1.386	1.886	2.920	4.303	6.965	9.925	31.598
3	0.765	0.978	1.250	1.638	2.353	3.182	4.541	5.841	12.941
4	0.741	0.941	1.190	1.533	2.132	2.776	3.747	4.604	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	3.365	4.032	6.856
10	0.700	0.879	1.093	1.372	1.812	2.228	2.764	3.169	4.587
20	0.687	0.860	1.064	1.325	1.725	2.086	2.528	2.845	3.850
30	0.683	0.854	1.055	1.310	1.697	2.042	2.457	2.750	3.646
40	0.681	0.851	1.050	1.303	1.684	2.021	2.423	2.704	3.551
60	0.679	0.848	1.046	1.296	1.671	2.000	2.390	2.660	3.460
120	0.677	0.845	1.041	1.289	1.658	1.980	2.358	2.617	3.373
∞	0.674	0.842	1.036	1.282	1.645	1.960	2.326	2.576	3.291

Sampling Error: (a probabilistic statement of sampling precision)

$$SE\% = \left(\frac{t_{\alpha,\, n-1}\ s_{\bar{x}}}{\bar{x}} \right) 100\%$$

The (1- α) sampling error is half the confidence interval expressed as a percentage of the mean.
It is a unit-less ratio and can be used to compare samples of different sizes and different units of measure. For example, your cannot statistically compare two sets of inventory results if one was computed with Doyle board feet and the other with cubic feet and each one had a different number of samples.
Since the standard error of the mean and the mean have the same unit of measure and were computed with n samples, the sampling error units of measure cancel and the t-value accounts for the probability associated with n samples. Thus, the sampling error can be used to compare the precision attained by two different samples.

Example of Confidence Interval and Sampling error

<u>Facts</u>: 10-1/5 acre plots per 40 acres that yielded $\bar{x}$=2,500 bd. ft/ac with s= ±250 bd. ft/ac

<u>Solution</u>: t @ 9 d. f. and α=0.05 ≐ 2.228

$$s_{\bar{x}} = \sqrt{\frac{250^2}{10}\left(1 - \frac{10}{200}\right)} = \pm 77.055 \text{ bd.ft} \qquad 95\%\text{ CI:} \quad 2{,}500 \pm (2.228_{05,\,9})(77.055) \qquad 2{,}500 \pm 171.68$$

$$SE\% = \left(\frac{171.68}{2{,}500} \right) 100\% \qquad = 6.9\% \text{ sampling error}$$

Sampling Error vs. Allowable Error

Sampling error is the precision achieved with the sampling procedure.
Allowable error is the desired/anticipated precision to be achieved with a sampling procedure.

Sample Size Required to a Achieve a Desired Precision

Sample Size: generic $$n = \frac{1}{\frac{1}{N} + \left(\frac{AE\%}{t\ CV\%}\right)^2} \qquad where\ N = \left(\frac{A}{ps}\right)$$

Sample Size: infinite population $$n = \frac{t^2 (CV\%)^2}{(AE\%)^2} = \left[\frac{tCV\%}{AE\%}\right]^2$$

Sample Size: finite population $$n = \frac{N\ t^2\ (CV\%)^2}{N\ (AE\%)^2 + t^2\ (CV\%)^2}$$

Sample size is relatively independent of tract/stand size.

Sample size is a function of the probability level to be achieved (i.e. t-value at α), variation in the population (i.e. s^2 as expressed by CV%), and the desired allowable error (i.e. half the anticipated confidence interval expressed as a percentage of the anticipated mean). CV% and AE% are percents; i.e. 60%, 10%.

The general sample size equation above collapses to the infinite when $N \rightarrow \infty$; i.e. $1/\infty = 0$.

Calculation of Sample Size:

1. Insert a starting value of t=1.96 (or even 2.0) for α=0.05 and ∞ degrees of freedom and calculate n.

2. Using the calculated value of n, choose an appropriate t-value at n-1 d.f. and recalculate n.

3. Repeat step 2 and the value of n should stabilize.

Example: (1) for a CV%=60%, AE%=±10%, and t=2; n= $\{2*60/10\}^2$ = 144
(2) t at 143 d.f. and α=0.05 = 1.98; thus $[1.98*60/10]^2$ = 141.1 = 142 samples

At a CV%=50% the sample size would be: $[1.98*50/10]^2$ = 99
At a CV%=60% the sample size would be: $[1.98*60/10]^2$ = 142
At a CV%=70% the sample size would be: $[1.98*70/10]^2$ = 192
This means that 142 sample would be needed for a 40 acre tract or a 4,000 acre tract; i.e. sample size is a function of variation and probability, not tract size.

Example of Sample Size computer program: <u>note that for a given CV and allowable error, the sample size does not vary substantially with tract size.</u>

NUMBER OF **.2 acre** SAMPLE POINTS for FINITE POPULATIONS @95%
with (number of points per 40 acres)

Tract Size	Coefficient of Variation	.......Allowable Error (percent of mean volume)...... 5%	10%	15%	20%	25%
40	50	132 (132)	66 (66)	36 (36)	22 (22)	15 (15)
80	50	198 (99)	79 (39)	39 (20)	23 (12)	15 (8)
120	50	237 (79)	84 (28)	41 (14)	24 (8)	15 (5)
160	50	263 (66)	87 (22)	41 (10)	24 (6)	15 (4)
200	50	282 (56)	89 (18)	42 (8)	24 (5)	15 (3)
400	50	328 (33)	93 (9)	43 (4)	24 (2)	16 (2)
600	50	347 (23)	95 (6)	43 (3)	24 (2)	16 (1)
640	50	349 (22)	95 (6)	43 (3)	24 (2)	16 (1)
40	60	148 (148)	83 (83)	48 (48)	30 (30)	20 (20)
80	60	234 (117)	104 (52)	54 (27)	32 (16)	21 (11)
120	60	291 (97)	114 (38)	57 (19)	33 (11)	22 (7)
160	60	331 (83)	120 (30)	58 (15)	34 (8)	22 (5)
200	60	361 (72)	124 (25)	59 (12)	34 (7)	22 (4)
400	60	440 (44)	132 (13)	61 (6)	35 (3)	22 (2)
600	60	475 (32)	135 (9)	61 (4)	35 (2)	22 (1)
640	60	480 (30)	135 (8)	62 (4)	35 (2)	22 (1)
40	70	159 (159)	98 (98)	60 (60)	39 (39)	27 (27)
80	70	263 (132)	130 (65)	70 (35)	43 (21)	29 (14)
120	70	337 (112)	146 (49)	75 (25)	44 (15)	29 (10)
160	70	392 (98)	155 (39)	77 (19)	45 (11)	30 (7)
200	70	435 (87)	161 (32)	79 (16)	46 (9)	30 (6)
400	70	555 (56)	175 (18)	82 (8)	47 (5)	30 (3)
600	70	612 (41)	181 (12)	83 (6)	47 (3)	30 (2)
640	70	620 (39)	181 (11)	83 (5)	47 (3)	30 (2)

n = [(N/PS)*(T^2.)*(CV%^2.)]/ [(N/PS)*(AE%^2.)+ (T^2.)*(CV%^2.)]
Assuming N is tract size in acres and average plot size **(PS) is .2 acres.**
T=Student's t (statistical value); approximately 1.96

NUMBER OF **0.05 acre** SAMPLE POINTS for FINITE POPULATIONS @95%
with (number of points per 40 acres)

Tract Size	Coefficient of Variation	.......Allowable Error (percent of mean volume)...... 5%	10%	15%	20%	25%
100	60	440 (176)	132 (53)	61 (24)	35 (14)	22 (9)
200	60	495 (99)	136 (27)	62 (12)	35 (7)	22 (4)
600	60	539 (36)	140 (9)	62 (4)	35 (2)	23 (2)
1000	60	549 (22)	140 (6)	63 (3)	35 (1)	23 (1)
1200	60	552 (18)	140 (5)	63 (2)	35 (1)	23 (1)
1400	60	554 (16)	140 (4)	63 (2)	35 (1)	23 (1)
1600	60	555 (14)	141 (4)	63 (2)	35 (1)	23 (1)

n = [(N/PS)*(T^2.)*(CV%^2.)]/ [(N/PS)*(AE%^2.)+ (T^2.)*(CV%^2.)]
Assuming N is tract size in acres and average plot size **(PS) is .05 acres.**