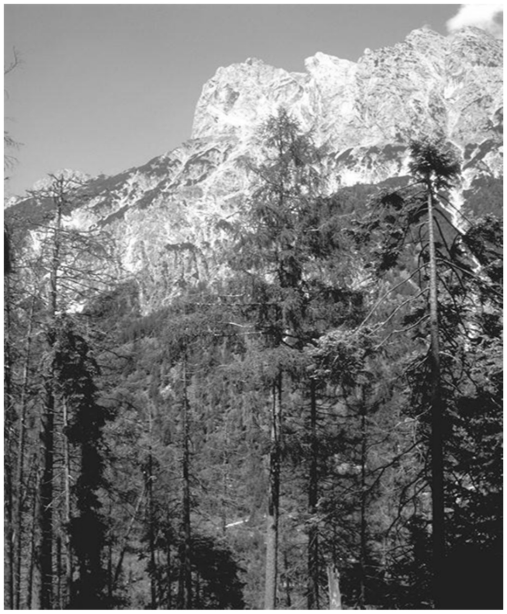


FORESTRY 1001
Fall 2011

Quote of the Day

Beer’s Hypothesis:

“*The most efficient sample design is one which samples proportionally to the variance of the parameter of interest.*”

- Tom Beers

Housekeeping Items

- Tuesday Nov 1 – MIDTERM!!!
  - Map Scale, Map Area, North Arrow
  - Height calculations from Suunto measurements
  - Mean, standard deviation, standard error
  - Stand and stock tables from tally data
  - Interpretation of Stand-level parameters
  - General familiarity with reading assignments in Forest Mensuration Book
- Tree ID Test Nov 10
  - Final Twig Key due

Stand Structure

- Size and spatial distribution of tree species
- Components
  - Where the trees are
  - What the trees are
  - How big the trees are

Stand Structure

- Overstory
- Understory
- Regeneration
- Forest Floor

Stand Structure

- Overstory

- Stand-Level Information
  - Species Composition
  - Size Distribution
    - Diameter
    - Height
    - Volume
  - Mean Tree Size
  - Crown Cover

- Individual Tree Information
  - Species
  - Stem Size
    - Diameter
    - Height
    - Volume or Weight
  - Crown Size
    - Length
    - Width

Populations

- The set of individuals of interest
- Individual/element – the fundamental unit of your population
  - Stands
  - Values per unit area
  - Individual Trees

Properties of Populations

- Individuals
- Countable (finite or infinite)
- Independent
- Often interested in certain parameters of the population
  - Total
  - Mean
  - Variation
  - Range
  - Distribution

Today's Population

Row # \ Column #	1	2	3	4
1	8	16	27	3
2	13	5	29	4
3	7	24	27	10
4	13	12	2	8

Properties of Populations

- Countable elements (finite or infinite)
- Independent elements
- True set of descriptive parameters
  - Mean
  - Variance

$$\mu = \frac{\sum_{i=1}^{N} X_i}{N}$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} (X_i - \mu)^2}{N}$$

Calculation of Population Parameters

- Census elements without measurement error
- Apply formulae

Population Mean

$$\mu = \frac{\sum_{i=1}^{16} X_i}{16}$$

$$= \frac{(8+16+2+13+5+29+4+3+7+24+27+104+13+12+2+8)}{16}$$

$$= \frac{208}{16} = 13\ \text{m}^3$$

Population Variance

$$\sigma^2 = \frac{\sum_{i=1}^{16} (X_i - 13)^2}{16}$$

$$= \frac{(8-13)^2 + (16-13)^2 + \cdots + (8-13)^2}{16}$$

$$== \frac{25+9+196+100+0+64+256+81+36+121+196+9+0+1+121+25}{16}$$

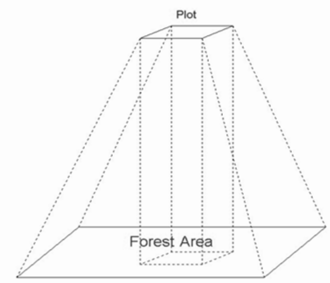
$$= \frac{1240}{16} = 77.5$$

Population Standard Deviation

$$\sigma = \sqrt{\sigma^2} = \sqrt{77.5} = 8.8$$

What is a sample?

- A subset, generally small, of the population of interest used to represent the whole population



Why Sample?

- Hard to measure without error
- Populations are large or even infinite
- Measurements are costly

Components of sample

- Sample Frame/List
- Finite countable set of elements
- Elements belong to the Population
- Elements are independent
- Elements are identically distributed
- Error

Sample Frame

- {1,1}
- {1,2}
- {1,3}
- {1,4}
- {2,1}
- {2,2}
- {2,3}
- {2,4}
- {3,1}
- {3,2}
- {3,3}
- {3,4}
- {4,1}
- {4,2}
- {4,3}
- {4,4}

Sampling

- Selecting a subset of elements from the sampling frame according to the rules of probability

Example sample of size 4

- Select 4 elements from our sample frame
- Generate 4 sets of random pairs {x,y} between 1 and 4
- {1,2}, {3,4}, {2,1}, {3,1}

Our Sample

Row # \ Column #	1	2	3	4
1	8	16	27	3
2	13	5	29	4
3	7	24	27	10
4	13	12	2	8

Our Data

- 16 m^3
- 10 m^3
- 13 m^3
- 7 m^3

Estimation of Population Mean

$$\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n}$$

$$\bar{x} = \frac{\sum_{i=1}^{4} X_i}{4}$$

$$= \frac{(16 + 10 + 7 + 13)}{4}$$

$$= \frac{46}{4} = 11.5 \text{ m}^3$$

Properties of Sample Mean

- Unbiased estimate of the population mean
- Maximum likelihood estimate of sample mean
- Sufficient estimator
- Other estimators:
  - Quadratic mean
  - Median
  - Mode

Estimation of Population Variance

$$s^2 = \frac{\sum_{i=1}^{n}(X_i - \bar{x})^2}{n-1}$$

$$s^2 = \frac{\sum_{i=1}^{4} X_i^2 - \frac{\left(\sum_{i=1}^{4} X_i\right)^2}{4}}{4-1}$$

$$s^2 = \frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{n-1}$$

$$= \frac{\left(16^2 + 10^2 + 7^2 + 13^2\right) - \frac{(16+10+7+13)^2}{4}}{3}$$

$$= \frac{574 - \frac{46^2}{4}}{3} = \frac{574-529}{3} = 15$$

Estimation of Population Standard Deviation

$$s = \sqrt{s^2} = \sqrt{\frac{\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}}{n-1}}$$

$$s = \sqrt{s^2} = \sqrt{15} = 3.87$$

Properties of Sample Variance

- Conditional – relies on estimate of mean
- Sufficient
- Unbiased
  - {n-1 recognizes that we used the data to estimate the mean}

Sampling Error

- Now we only selected 4 of 16 elements from the population
- Our estimates have error associated with them
- If we sample another 4 elements, we most likely will get different answers
- We need to assess that error

Sampling Error

- Now we only selected 4 of 16 elements from the population
- Our estimates have error associated with them
- If we sample another 4 elements, we most likely will get different answers
- We need to assess that error

How many samples of size 4 are there available? (assume no replacement)

$$C = \frac{N!}{n!\left(N-n\right)!}$$

$$\begin{aligned} C &= \frac{16!}{4!\left(16-4\right)!} \\ &= \frac{16!}{4! \cdot 12!} \\ &= \frac{16 \cdot 15 \cdot 14 \cdots 3 \cdot 2 \cdot 1}{\left(4 \cdot 3 \cdot 2 \cdot 1\right)\left(12 \cdot 11 \cdot 10 \cdots 3 \cdot 2 \cdot 1\right)} \\ &= \frac{16 \cdot 15 \cdot 14 \cdot 13}{4 \cdot 3 \cdot 2 \cdot 1} \\ &= \frac{43680}{24} \\ &= 1820. \end{aligned}$$

Sampling Error

- Measure the difference between an estimated mean and the true population mean
- Calculate the standard deviation of the means

MEAN of means

$$\bar{\bar{x}} = \frac{\sum_{j=1}^{m} \bar{x}_j}{m}$$

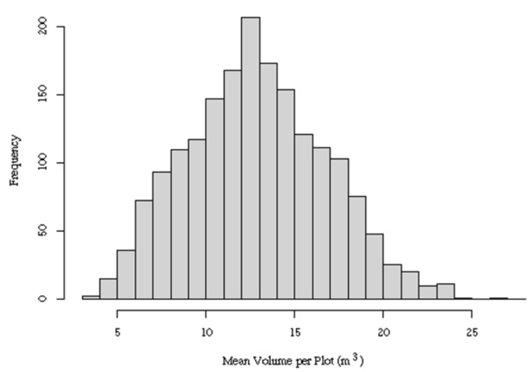
where

$$\bar{x}_j = \frac{\sum_{i=1}^{n} X_{i,j}}{n}$$

Standard deviation of means

$$s_{\bar{\mathbf{x}}} = \sqrt{\frac{\sum_{\mathbf{j}=1}^{\mathbf{m}} \left(\bar{\mathbf{x}}_{\mathbf{j}} - \bar{\bar{\mathbf{x}}}\right)^2}{\mathbf{m} - 1}}$$

Distribution of Means of size 4



[Figure 3-2]

Estimation of Sampling Error

- Previous formula requires repeated sampling
- Costly and inefficient
- Central Limit Theorem
  - Early sample experiments noted some common features of the distribution of means
  - Shape of distribution was often bell-shaped regardless of the underlying population distribution

Central Limit Theorem

- Regardless of the underlying population distribution the sample distribution (ie, the distribution of the means) approaches a Normal distribution with mean μ and variance $\sigma/\sqrt{n}$

Normal Distribution

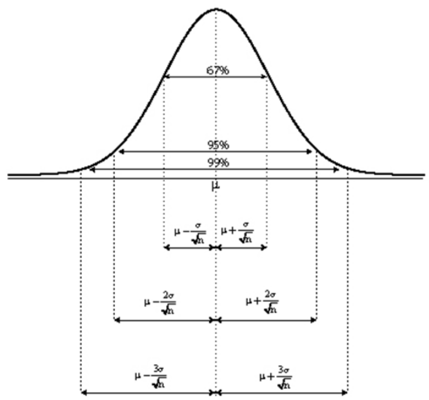


Figure 3-4

Results of the CLT

- Distribution of means is known (approximately)
- Estimate sampling error from a single sample:

$$s_{\bar{x}} = \frac{s}{\sqrt{n}}$$

- Standard Error of Estimate (Mean)

Some properties of Standard Error

- Unbiased estimate of standard deviation of means for small sample sizes
- Estimate -> 0 as n -> ∝
- Over estimates error (biased estimate) when sample size is large

Consider our census

- Based on population

$$s_{\bar{x}} = 8.8 / \sqrt{16} = 2.2$$

- Finite Correction Factor

$$s_{\bar{x}} = \frac{s}{\sqrt{n}} \cdot \sqrt{\frac{N-n}{N}}$$

Our Sample of Size 4

- Uncorrected

$$s_{\bar{x}} = \frac{3.87}{\sqrt{4}} = \frac{3.87}{2} = 1.94$$

- Corrected

$$s_{\bar{x}} = \frac{3.87}{\sqrt{4}} \cdot \sqrt{\frac{16-4}{16}} = 1.94 \cdot \sqrt{\frac{12}{16}} = 1.94 \cdot 0.866 = 1.68$$

Confidence

- Confidence allows us to bound our sample estimates with a given probability
- We make inferences about what future samples would look like based on current sample
- Allows us to measure precision

Confidence

- t-distribution gives values for a sample distribution with mean = 0 and variance = 1 for a given sample size
- We use this “standard” distribution to bound our observed distribution
- So we convert this t(0,1) into a t(mean,variance)

The distribution again

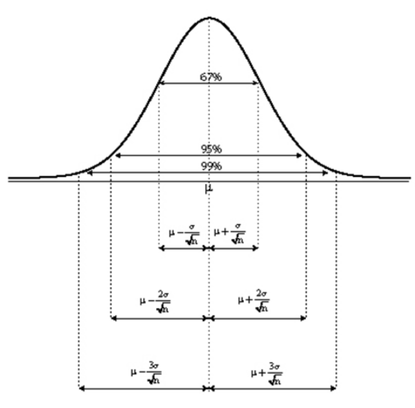


Figure 3-4

Confidence Interval

$$\text{C.I.} = \bar{x} \pm t \cdot s_{\bar{x}}$$

95% Confidence Interval for our Sample

- n = 4, CI = .95
- T((1-.95),4-1) = 3.182
- CI: $$\begin{aligned} \text{C.I.} &= 11.5 \pm t \cdot 1.94 \\ &= 11.5 \pm 3.182 \cdot 1.94 \\ &= 11.5 \pm 6.16 \end{aligned}$$
- CI: Pr(5.3 ≤ μ ≤ 17.7) = .95

Minimum Sample Size

Systematic Sampling

- samples laid out along a systematic grid
- logistically the easiest
- unbiased mean
- biased standard deviation

Systematic Sample of our population

Column # / Row #	1	2	3	4
1	8	16	27	3
2	13	5	29	4
3	7	24	27	10
4	13	12	2	8

Sampling Frame?

Questions

- How many independent systematic samples of size 4 are there?
- For the sample illustrated, how many independent choices were made?

Why a Biased Standard Deviation?

- Consider the following example:
  - We measure a single sample plot (ie, 1 sample unit)
  - We measure the dbh of every tree on the sampling unit
- How many individual choices of sampling units did you make?
- Can we calculate the standard error for mean basal area per hectare?
  - Why?
- Can we calculate the standard error for mean dbh?
  - Why?
  - Any limits to this estimate?

Why a Biased Standard Deviation?

- Now consider a systematic sample of 4 sample plots
  - Plots are located 200 m apart on a 2 by 2 grid
  - Start point randomly selected
- How many individual choices did you make?
- You have 4 sample plots, so you "CAN" calculate a standard error of mean basal area per ha, but what is it analogous to?

All Systematic samples of size 4

Row # \ Column #	1	2	3	4
1	8	16	27	3
2	13	5	29	4
3	7	24	27	10
4	13	12	2	8

Properties of systematic samples

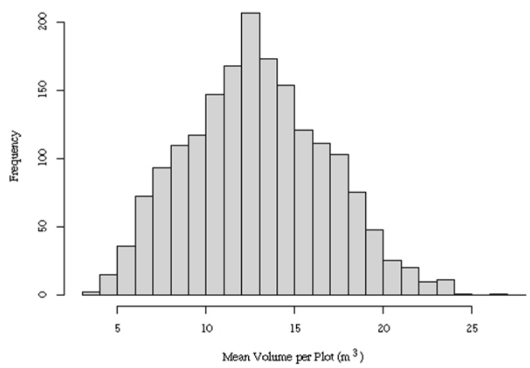
- Unbiased mean
- One sample selection (once 1 plot is located, all others are fixed)
- Sampling units are not independent
- Need at least two independent selections to calculate unbiased estimate of standard deviation (remember the n-1)
- s, as estimated from systematic sample, tends to be too large

Why is s too large?

- How many random samples of size 4?
- What is range of random samples of size 4?
- How many systematic samples of size 4?
- What is range of systematic samples of size 4?
- So what does systematic sampling do for us?

Distribution of Means of size 4

1820 samples
Possible!



Range of means: 3.5 – 26.75

[Figure 3-2]

Systematic Sample of Size 4

- Only 4 samples possible
  - 7.25, 13.25, 16.25, 17.25
- Clearly not as much variance possible
- "True" standard error would be the standard deviation of these 4 means

So Why use Systematic Sampling?