http://www.gov.mb.ca/conservation/forestry/renewal/surveys.html#ftgmethods

Sampling Concepts
- Part I -

FOR 1001
Dr. Thom Erdle

Objectives – Sampling Concepts

- **Introduce key concepts about *statistical inference***
- **Understand *relationships* between:**
  - ***variability***
  - ***confidence***
  - ***probability***
  - ***sample size***
- **Calculate *confidence interval* for an estimate**
- **Determine a *sample size* to provide estimate that meets *desired level of confidence***

Topics

- Populations & Parameters
- Measures of Central Tendency
- Sampling & Sampling Strategy Elements
- Calculating Confidence Interval
- Determining Sample Size

Population Parameters

<u>Population</u> – *what is it?*

- the entire group of "individuals" of a specific category within an area of interest
- defined by the context of the problem or issue in question

e.g.
- stands in Noonan forest
- shade trees on UNB campus
- salmon in Miramichi River
- forestry/ENR students at UNB

Population Parameters

<u>Parameter</u> – *what is it?*

- a characteristic of the population
- governed by the distribution of values across members of the population

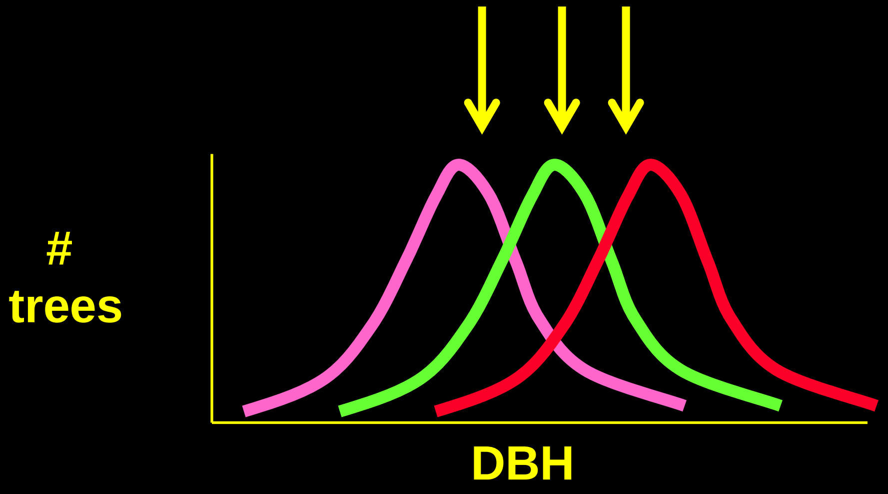
e.g.
- wood volume in stands in Noonan Forest
- # shade trees on UNB campus
- weight of fish in Miramichi River
- summer employment income of forestry /ENR students

Population Parameters

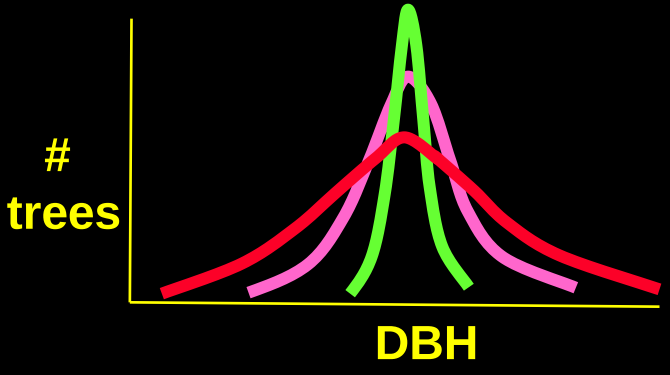
Parameter

- often characterized by Measures of Central Tendency
- relate to distribution of values in population elements

"Central" Value



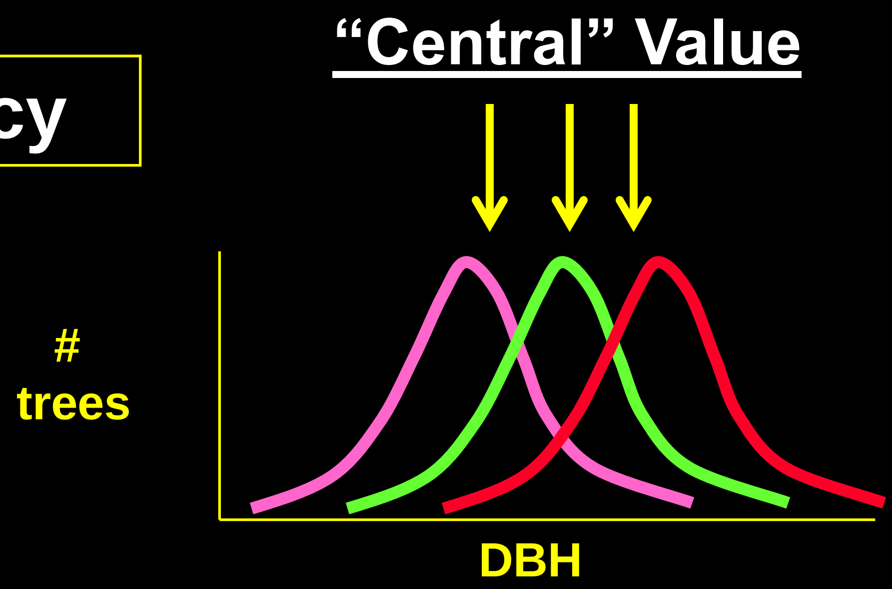
Spread



Topics

- Populations & Parameters
- *Measures of Central Tendency*
- Sampling & Sampling Strategy Elements
- Calculating Confidence Interval
- Determining Sample Size

Measures of Central Tendency



Central Values

Mean = $\dfrac{\text{sum values for each element}}{\text{number of elements}}$

Mode = most frequently occurring value

Median = half of values fall above; half fall below

Measures of Central Tendency

Central Values

Mean = ∑ values / # elements
= 280 / 11
= 25.5 cm

Mode = most frequent value
= 22 cm

Median = midpoint value
= half values above; half below
= 24 cm

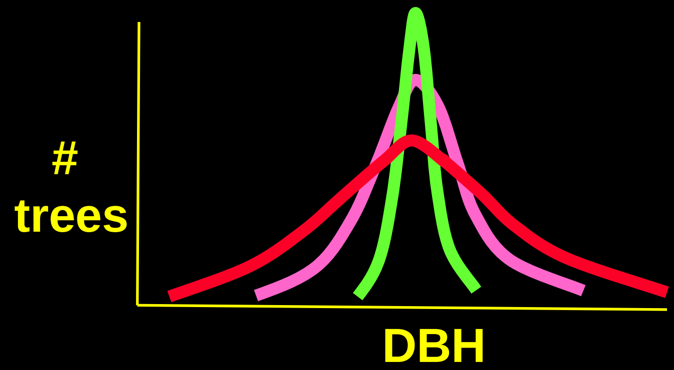
Tree	DBH_{cm}
1.	20
2.	22
3.	22
4.	22
5.	22
6.	24
7.	24
8.	26
9.	28
10.	34
11.	36

∑ = 280

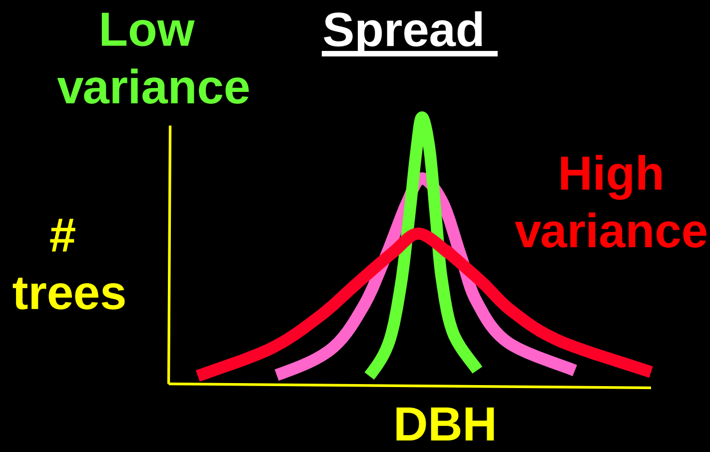
Measures of Central Tendency

Spread …relating to variation of values within population

Spread



Measures of Central Tendency



Spread

Variance = average of squared differences from mean value

Variance = $\sum (Y_i - Y_{mean})^2/(n - 1)$

Standard Deviation = SQRT(Variance)

DBH_{cm}	$(Y_i - Y_{mean})^2$
20	$(20-24)^2 = 16$
22	$(22-24)^2 = 4$
24	$(24-24)^2 = 0$
26	$(26-24)^2 = 4$
28	$(28-24)^2 = 16$

$\sum = 120$
Mean = 120/5
Mean = 24cm

$\sum = 40$
Var = 40 / 4
Var = $10cm^2$
StDev = 3.16cm

Topics

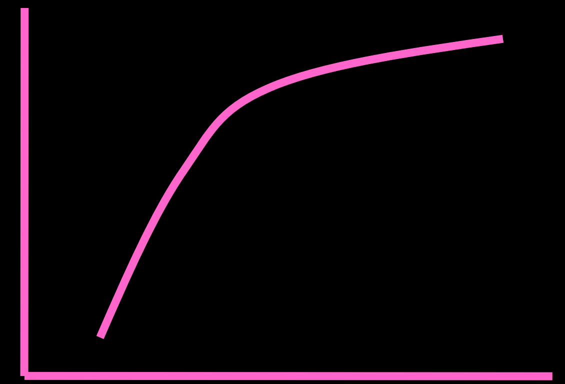
- Populations & Parameters
- Measures of Central Tendency
- ***Sampling & Sampling Strategy Elements***
- Calculating Confidence Interval
- Determining Sample Size

Sampling

If we want to know something about population parameters, what to do?

Census	Sample
Observe *each* element in population	Observe *some* elements in population
Maximum confidence in findings	Confidence = *f{sampling strategy,variation}*
Zero sampling error	*Some* sampling error
May be *expensive* & time consuming	Can be *cheaper* & less time consuming
Representativeness *guaranteed*	Representativeness *function of sampling* design
No inference necessary	*Inference* necessary

Sampling



- **Why in forestry & natural res?**
  - forest/enviroment is big
  - field work is costly
  - access can be difficult (in some places)
  - diminishing returns
- **So, sampling *frequently used* in forestry & natural resources to estimate population parameters**
- **Need *well-designed* sampling *strategy* to generate desired results**
- ***Low cost* is no virtue *without quality***

Sampling

- **Sampling Strategy**
  - governs **quality** of population parameter **estimates**
  - ***central concern*** of forest management
  - requires ***statistical know-how*** & familiarity with ***basic concepts***

Sampling Strategy

What are the elements?

- Choice of *sampling units*
  - type of plot
  - size of plot
- *Allocation* of samples across population of interest
  - random
  - systematic
  - stratified
- *Intensity* of sampling
  - number of plots (samples)
- *Observations* to record
  - what measurements to take & how
- *Timing* & other considerations

Next few weeks

Our concern today

Sampling Estimate

- **Sampling can only *estimate* the population parameter**
  - so, how good is the estimate?
  - or, how can we make the estimate good?
  - what governs quality of estimate?

- **The *anatomy* of a sample estimate**
  - population variability
  - confidence interval
  - probability
  - sample size

Sampling Estimate

An example – something you are likely to hear or read

"The Pulse Market Research Agency has recently conducted a poll of Canadian voters and found that…."

- ***35%*** of the ***1150 people*** sampled support the Conservative Party
- accuracy is ***3% 19 times*** out of ***20***

sample size

estimate

Confidence limit

probability

Sampling Estimate

An example – something you are likely to produce

"We have completed our survey of the Noonan forest and estimate that it contains…."

- 215 m^3/ha based on 8 samples (1/10th ha each)
- we are 95% confident that the true value lies between 200 & 230 m^3/ha

sample size

estimate

probability

Confidence interval

Sampling Estimate

❑ Confidence Interval

- each estimate has an associated ***confidence interval***
- the estimate the ***confidence limits*** gives the ***confidence interval***
- the ***true*** population parameter is deemed to ***fall within*** this ***interval*** with a certain ***probability***

❑ Probability level

- the probability that the ***true*** population ***value falls within*** the confidence ***interval***

Sampling Estimate

I estimate the world's population to be……

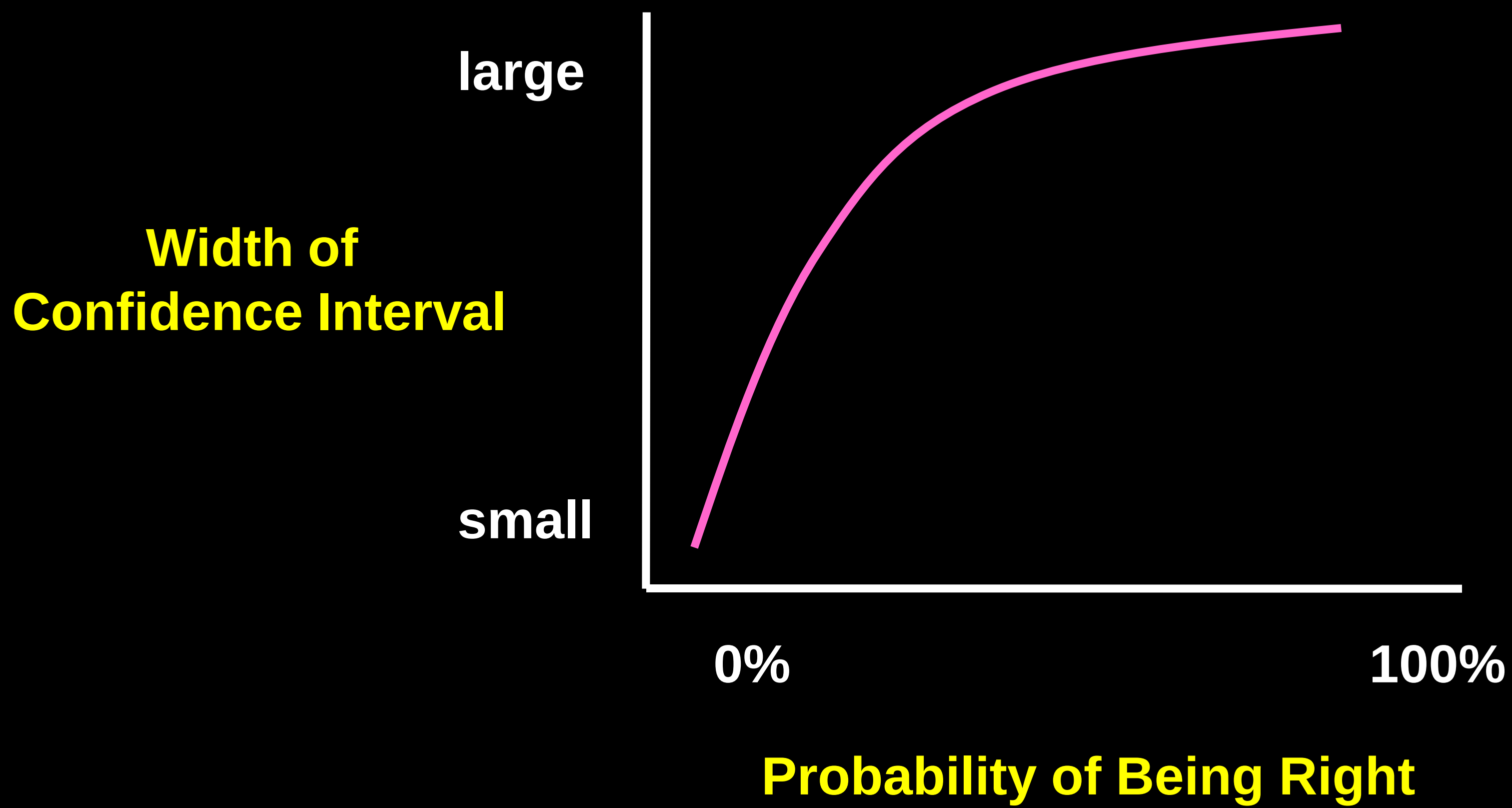
Confidence Interval	Probability level
Between *0* and *10* billion	Almost certain to be *true(100%*)
Between *2* and *8* billion	*99%* probability of being true
Between *5* and *7* billion	*98%* probability of being true
Between *5.5* and *7.5* billion	*90%* probability of being true
Between *6.9* and *7.1* billion	*75%* probablity of being true
Between *6,999,999,999* and *7,000,000,001* billion	Virtually *0%*

Sampling Estimate

I estimate the m^3/ha in Noonan stand #2021 to be…

Confidence Interval	**Probability level**
Between *0* and *500*	**Almost certain to be *true(100%*)**
Between *100* and *400*	***99%* probability of being true**
Between *150* and *350*	***90%* probability of being true**
Between *190* and *210*	***30%* probability of being true**
Between *214*and *215*	***<1%* probablity of being true**

Sampling Estimate



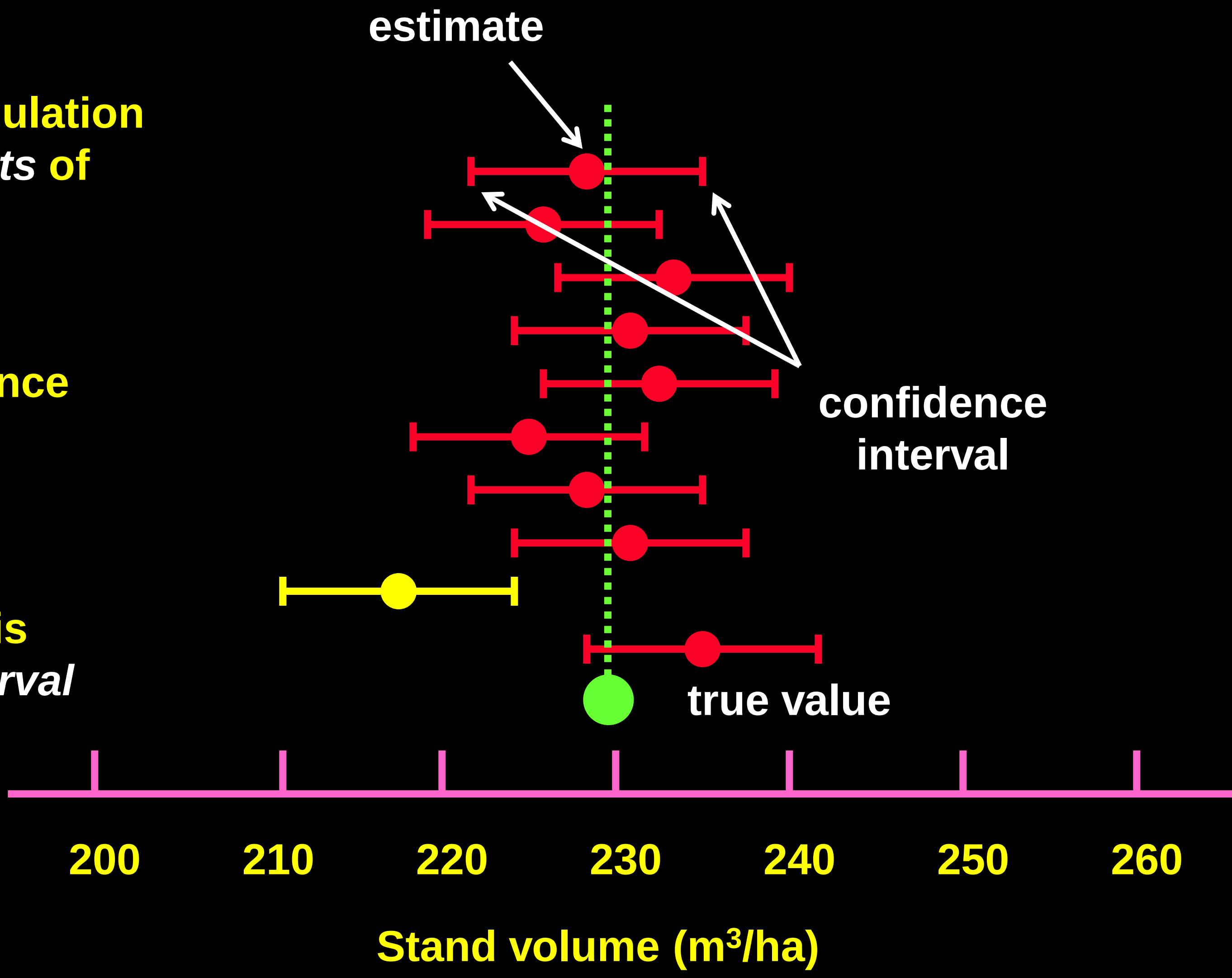
Probability Level

- **Example : 90% Probability**

If we were to estimate the population parameter with *10 different sets of* samples…..

…. the *true parameter* value would *fall outside* the confidence *interval* in *1* of those *10 cases*

There is a *one-in-ten* chance that the *true* parameter *value* is *not within* the *confidence interval*



Topics

- Populations & Parameters
- Measures of Central Tendency
- Sampling & Sampling Strategy Elements
- *Calculating Confidence Interval*
- Determining Sample Size

Example

- We want to *estimate* the *density* (stems/ha) in a stand
- We establish *five 0.1ha plots* and *count trees* in each

Plot #	Tree Count (stems/plot)	Density (stems/ha)
1	40	400
2	27	270
3	23	230
4	47	470
5	46	460
Avg		366

- We *estimate* the density to be *366 stems/ha*
- How *good* is it?

= 366 stems/ha ± 1

= 366 stems/ha ± 10

= 366 stems/ha ± 100

= 366 stems/ha ± 300

Confidence Interval

- ***Bounds*** **placed on estimate**
- **Function of "*standard error*"**

Standard error = Standard Deviation / sqrt (sample size)

Standard Error = SQRT [s^2/n (1 - n/N)]

Where:	**s**	**=**	**standard deviation**
	1 - n/N	**=**	**finite population correction**
	N	**=**	**population size**
	n	**=**	**sample size**

But n/N is usually small so generally omit

Standard Error = SQRT (s^2/n) = s / sqrt(n)

Confidence Interval

- **Standard Error Example**

Variance = $\sum (400-366)^2 + (270-366)^2 \ldots) / (5-1)$

Plot #	Tree Count (stems/plot)	Density (stems/ha)
1	40	400
2	27	270
3	23	230
4	47	470
5	46	460
Avg		366

Variance = 12130 $(stems/ha)^2$

Standard Deviation = 110.1 stems/ha

Standard error of mean

= 110.1 / sqrt(5) = 49.2 stems/ha

Standard error of mean (what influences it?)

- decreases as variation decreases
- decreases as # samples increases

Standard Error = SQRT (s^2/n) = s / sqrt(n)

Confidence Interval

❑ Confidence Interval

Confidence interval = Mean ± Confidence limits

Confidence interval = mean ± Standard Error * t $_{\text{(@ chosen probability \& n-1 df)}}$

❑ Confidence Limits

Confidence limits = Standard Error * t $_{\text{(@ chosen probability \& n-1 df)}}$

❑ t - value

Function of: chosen probability
degrees of freedom (sample size minus 1)
draw from statistical 't' table

Confidence Interval

- **Standard Error Example**

Variance = $\sum (400-366)^2 + (270-366)^2 \ldots) / (5 - 1)$

Plot #	Tree Count (stems/plot)	Density (stems/ha)
1	40	400
2	27	270
3	23	230
4	47	470
5	46	460
Avg		366

Variance = 12130 (stems/ha)^2

Standard Deviation = 110.1 stems/ha

Standard error of mean

= 110.1 / sqrt(5) = 49.2 stems/ha

Standard error of mean

- decreases as variation decreases
- decreases as # samples increases

Confidence Interval

- ## Confidence Interval

  Confidence interval = Mean ± Confidence limits

  Estimate of Density = 366 stems/ha ± Confidence limits

  Estimate of Density = 366 stems/ha ± 137 = 229 to 503 stems/ha

- ## Confidence Limit

  Confidence limit = Standard Error * $t_{(@\ \text{chosen probability}\ \&\ n-1\ \text{df})}$

  Confidence limit = 49.2 * $t_{(95\%\ \text{probability}\ \&\ 4\ \text{df})}$

  Confidence limit = 49.2 * 2.78

  Confidence limit = 137 stems/ha

- ## t - value

  ‘t-value’ for 95% probability & 4 degrees of freedom = 2.78

Confidence Interval

Plot #	Tree Count (stems/plot)	Density (stems/ha)
1	40	400
2	27	270
3	23	230
4	47	470
5	46	460
Avg		366

- **Interpretation of findings**
- **From these sample results we can say…..**
- **We are *95% confident* that that *true density* in the sampled stand is between *229* and *503 stems per ha***
- **There is only *one chance out of 20* (thus 95% confidence) that the *true density* in the sampled stand is between *229 and 503 stems per ha***

Confidence Interval

- **What happens to confidence interval when *variation*, *probability level*, and *sample size* change?**