FOR 474: Forest Inventory Techniques

Inventory Estimation

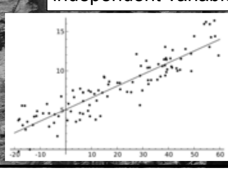
- Regressions
- Regression Estimation
- Ratio Estimation
- Mean of Ratios Estimation
- PPP Sampling

Avery and Burkhart Chapters 8 & 12 (4th edition)
http://www.fs.fed.us/fmsc/measure/handbooks/index.shtml

Regression: What is it?

The statistical technique for estimating the relationship among variables.

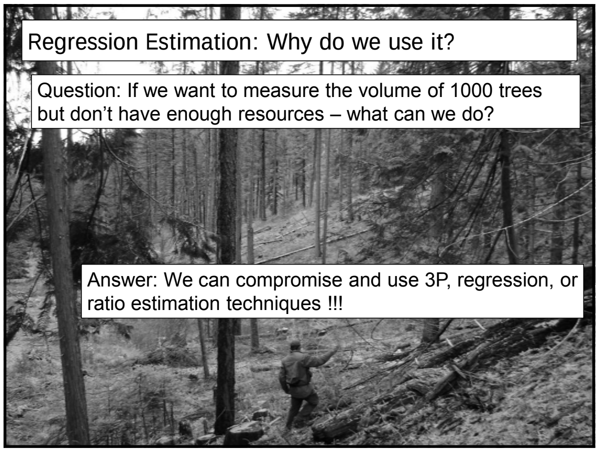
The premise:
1. Collect one or more independent variables (easy to measure)
2. Determine their relationship to the dependent variable (difficult to measure), determining how changes in the independent variables influence the dependent variable.

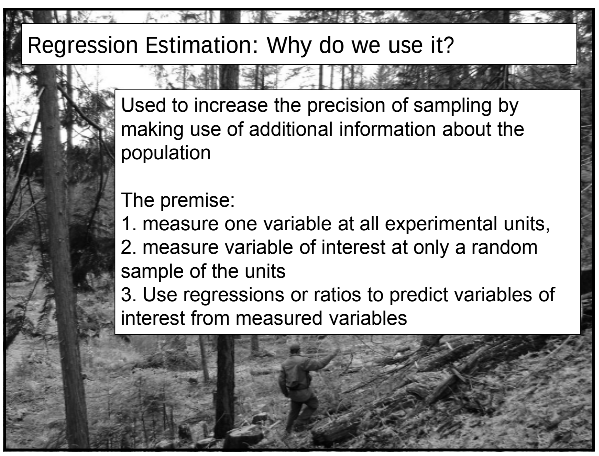


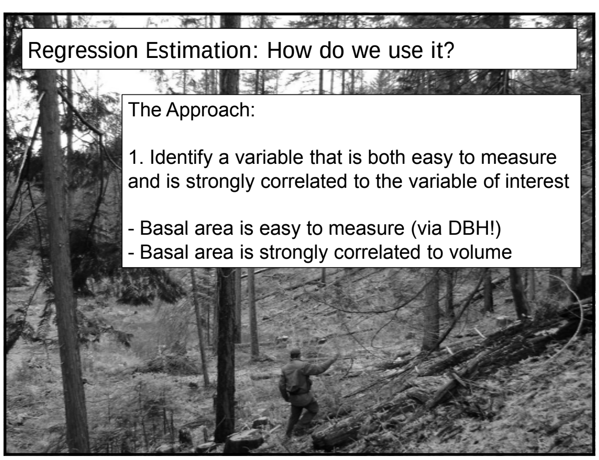
Regression: Confidence in the relationship

Commonly assessed by the coefficient of determination (r^2).

The premise:
1. Describes how well a regression line fits a set of data, scaled from 0 to 1.
2. The closer to 1, the more variability the independent variables are able to explain about the dependent variable.
3. Provides a measure of how well future outcomes are likely to be predicted by the model.


Regression Estimation: Why do we use it?
Question: If we want to measure the volume of 1000 trees but don't have enough resources – what can we do?
Answer: We can compromise and use 3P, regression, or ratio estimation techniques !!!


Regression Estimation: Why do we use it?
Used to increase the precision of sampling by making use of additional information about the population
The premise:
1. measure one variable at all experimental units,
2. measure variable of interest at only a random sample of the units
3. Use regressions or ratios to predict variables of interest from measured variables


Regression Estimation: How do we use it?
The Approach:
1. Identify a variable that is both easy to measure and is strongly correlated to the variable of interest
- Basal area is easy to measure (via DBH!)
- Basal area is strongly correlated to volume

Regression Estimation: How do we use it?

The Approach:

2. Know the population mean or total of that second variable (mean basal area)

Example: We want to measure volume of 1,273 trees but don't have the resources to do it

- Measure basal area of all 1,273 trees
- Basal area population mean, μ_x = 1.40 sq ft

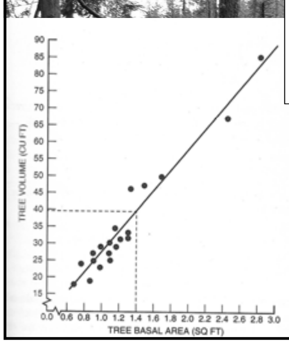
Regression Estimation: How do we use it?

The Approach:

3. Measure simple random sample of both variables

From simple random sample of 20 trees:
Basal area sample mean, $\bar{x}$ = 1.27
Volume sample mean, $\bar{v}$ = 35.1 (ft^3)

Regression Estimation: How do we use it?



The Approach:

4. Plot data between sample volume and basal area to calculate gradient of straight line

$$b = \frac{V_2 - V_1}{x_2 - x_1}$$

$b = regression\ gradient$

Example: b = 30.65

Regression Estimation: How do we use it?

The Approach:

5. Calculate regression to combine datasets

$$V_R = \bar{v} + b(\mu_x - \bar{x})$$

$V = volumes$

$\bar{x} = sample\ mean\ of\ basal\ areas$

$\mu_x = population\ mean\ of\ basal\ area$

$b = regression\ gradient$

V_R = 35.1 + 30.65*(1.40-1.27) = 39.1 (ft^3)

Regression Estimation: When can we use it?

Several conditions have to be met to use regression estimation:

1. Need to know population mean for x
2. Need mostly linear relationship
3. Variance of y about mean should be same for all values of x

However, if we know how the variance of y changes with x we can use a “weighted regression”

Regression Estimation: Ratio of Means

A Special Case:

- Regression is a straight line through the origin (so no intercept)
- Standard deviation in y proportional to sq root of x

$$y_R = RX = \frac{\bar{y}}{\bar{x}} X = \frac{\sum y}{\sum x} X$$

X = Known population mean of x

Regression Estimation: Ratio of Means

Example:

Population N = 400
Population mean of x (DBH) = 11.2
Following n = 10 sample:

	DBH	Height		
	8	62		
	13	81		
	5	40		
	6	46		
	19	123		
	9	74		
	8	52		
	11	96		
	5	36		
	12	70		
Total	96	680	R =	7.083
Mean	9.6	68		

y = RX
y = 7.083 * 11.2 = 79.3

Regression Estimation: Mean of Ratios

A Special Case:

- Regression is a straight line through the origin (so no intercept)
- Standard deviation in y proportional to x (other than sq root)

$$y_R = \hat{R}X = \frac{\sum r}{n} X$$

$\hat{R} = \text{mean of individual ratios}$

X = Known population mean of x

Regression Estimation: Mean of Ratios

Example:

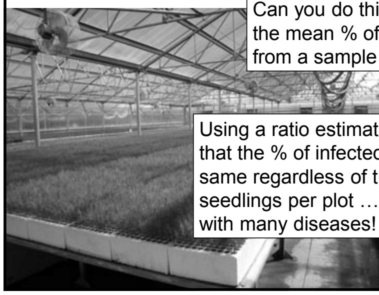
Population N = 100
Population mean of x = 76
Following n = 10 sample:

x	y	r
36	18	0.500
95	48	0.505
108	46	0.426
172	74	0.430
126	58	0.460
58	26	0.448
123	60	0.488
98	51	0.520
54	25	0.463
14	7	0.500
	Total	4.741
	Mean	0.474

y = RX
y = 0.474 * 76 = 36.024

Regression Estimation: Food for thought …

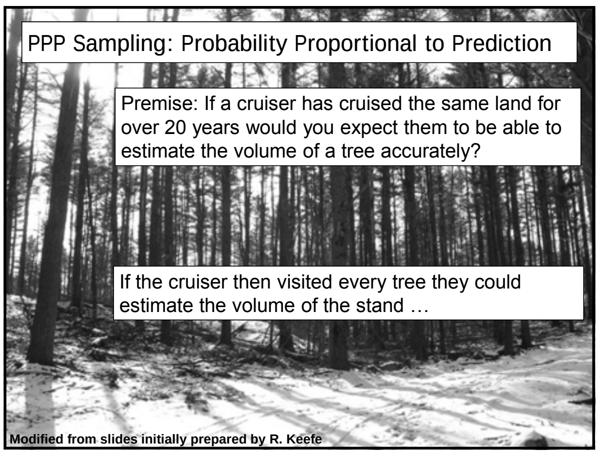
Imagine you want to estimate the total number of infected seedlings in the University nursery …



Can you do this by just multiplying the mean % of infected seedlings from a sample by the total number?

Using a ratio estimator we would assume that the % of infected seedlings is the same regardless of the number of seedlings per plot … this is <u>clearly not true</u> with many diseases!

PPP Sampling: Probability Proportional to Prediction

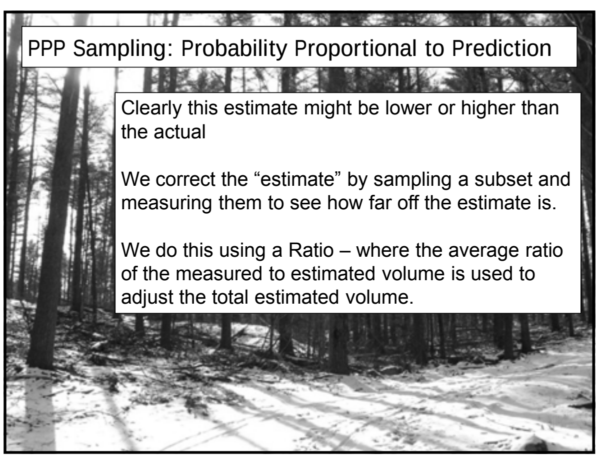


Premise: If a cruiser has cruised the same land for over 20 years would you expect them to be able to estimate the volume of a tree accurately?

If the cruiser then visited every tree they could estimate the volume of the stand …

Modified from slides initially prepared by R. Keefe

PPP Sampling: Probability Proportional to Prediction



Clearly this estimate might be lower or higher than the actual

We correct the “estimate” by sampling a subset and measuring them to see how far off the estimate is.

We do this using a Ratio – where the average ratio of the measured to estimated volume is used to adjust the total estimated volume.

PPP Sampling: Some Notes from the Designer

Adv - "SPEEDY OCULAR ASSESSMENT OF SOME QUANTITY ASSOCIATED WITH AN INDIVIDUAL TREE AT THE TIME IT IS VISITED AND MARKED FOR HARVEST IS CHEAP…"

Adv - "TALENTED ASSESSORS CAN QUICKLY MAKE BIASED OCULAR JUDGMENTS AT THE TIME OF MARKING THAT WILL BE HIGHLY CORRELATED WITH INDIVIDUAL TREE QUANTITIES IMPOSSIBLE TO MEASURE IN THE STANDING TREE …"

Dis - "CORRECTION OF SUCH BIASED JUDGMENT … INVOLVING REGRESSION OR RATIO ESTIMATES HAS BEEN UNSATISFACTORY BECAUSE OF ERRATIC VARIATION AND UNKNOWN FORM OF UNDERLYING FUNCTIONAL RELATIONSHIPS…"

L.R. Grosenbaugh (1965)

PPP Sampling: Probability Proportional to Prediction

The Steps:

1. Visit all trees and *predict* attribute of interest
2. For subset, also *measure* attribute of interest
3. Correct overall estimate based on ratio
4. <u>Add anything above/outside original range</u>

Note:

The attribute does not have to be volume/value – it could be anything you are interested in

PPP Sampling: When do we use PPP

When to use:

- Precise estimates are sought
- High value timber. E.g., nice hardwoods
- Small # of stems? Not necessarily…
- Used in NIPF sales where marking was based on complex management objectives, but then by the time of sale – value is wanted

Note: not a feasible alternative to fixed area or point sampling methods is often used as a quick way to estimate volume in "lump-sum" timber sales

NIPF = non-industrial private forest

PPP Sampling: Before We Head out

1. Use the preliminary sample size equation to determine the number of trees you need to measure

$n = t^2CV^2/\%E^2$

n = samples of trees to achieve precision of allowable percentage error, %E (for PPP typically set to 1.5%)
t = t-value (2 for 95% level)
CV = coefficient of variation of y_i
y_i= ratio of measured volume to estimated volume

Note: Most cruisers get CV to be within 15%

PPP Sampling: The Estimation Steps

2. Estimate the total stand tree volume, T, based on all the cruiser's estimates (or guesses! = $X_1, X_2, X_3, \ldots. X_n$)
3. Estimate the maximum volume that an individual tree can have, K (i.e. K = maximum X_i volume)
4. Estimate the total number of trees in your stand, N
5. Next generate a set of random numbers of size N (i.e. how many trees you will visit) of values from 1 to K (i.e. based on the volumes and not the number of trees!)

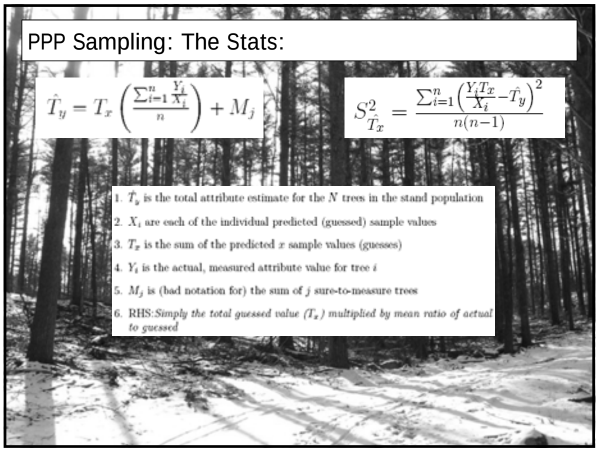
Assign in a list, a random volume number to each tree that you will visit in your cruise

PPP Sampling: The Measurement Steps

5. Next 2 cruisers head out and the 1st guesses volume of each tree and records the estimates
6. 2nd cruises compared guess to the random number (of volume) that is assigned to that tree number they are on

   If estimate > random number assigned to that tree:
   Tree is measured in detail

   If any tree has an estimate > K (i.e. the believed maximum volume) then it is also measures


PPP Sampling: The Stats:
$\hat{T}_y = T_x \left(\frac{\sum_{i=1}^{n} \frac{Y_i}{X_i}}{n} \right) + M_j$
$S^2_{\hat{T}_x} = \frac{\sum_{i=1}^{n} \left(\frac{Y_i T_x}{X_i} - \hat{T}_y \right)^2}{n(n-1)}$
1. $\hat{T}_y$ is the total attribute estimate for the N trees in the stand population
2. X_i are each of the individual predicted (guessed) sample values
3. T_x is the sum of the predicted x sample values (guesses)
4. Y_i is the actual, measured attribute value for tree i
5. M_j is (bad notation for) the sum of j sure-to-measure trees
6. RHS: *Simply the total guessed value (T_x) multiplied by mean ratio of actual to guessed*