

# ESTIMATING PROPORTION OF BIRCH TREES IN PLOT SOUTH OF REICHARDT

STEFANO FOCHESSATTO, BRIA HIEBERT-CRAPE, THOMAS HOUSE

**ABSTRACT.** Proportion data was collected from the plot directly south of the Reichardt building at UAF during the Fall 2021 semester. The plot was organized into 173 north to south transects to facilitate an arms-length sample width. The entire population of trees in each transect was then sampled, thus performing a one-stage cluster sample. Our group decided on taking 6 samples, two for each sampler. The proportions estimate was computed using the usual one-stage cluster sample proportion estimator.

## AREA STUDY

**Initial Survey and Area Analysis.** With satellite imagery we found that the full length of the plot as measured on the northernmost side came out to be 865 feet. An initial survey found the proportion of birch trees in the plot to be very low and relatively clustered together. The plot contained mainly three species of trees: spruce, aspen, and birch.

The transects were plotted north to south to span all elevations, despite being more difficult to sample. We decided to make each transect about arms-length to make it easier to sample every tree in the transect. With the width of each transect and the length of the northernmost side we can compute the total number of transects(clusters) through division.

## 1. SAMPLE ALLOCATION AND METHODS

Our group decided to sample only 6 samples and divide them evenly among each group member. Since we are using single stage cluster sampling we can always sample more clusters to improve our estimate. With the total number of transects computed ( $N$ ) and the total number of transect samples ( $n$ ) decided, we computed a simple random sample of the transects in  $r$  to decide which transects to sample.

FIGURE 1. Sample Allocation Code:

```

#SRS of clusters - 5' wide
#Length of northernmost side of plot
Length = 865

#Population of Clusters with width 5.
N <- Length/5
n <- 6

#Setting seed and applying srs to assign clusters.
set.seed(102221); samp <- sample(N, n, replace = F)
samp_assigned <- matrix(samp, nrow = 3, byrow = T,
                        dimnames = list(c("Stefano", "Thomas", "Bria"),
                                      c("samp1", "samp2")))
> samp_assigned
# samp1 samp2
#Stefano 107 158
#Thomas 145 91
#Bria 19 130

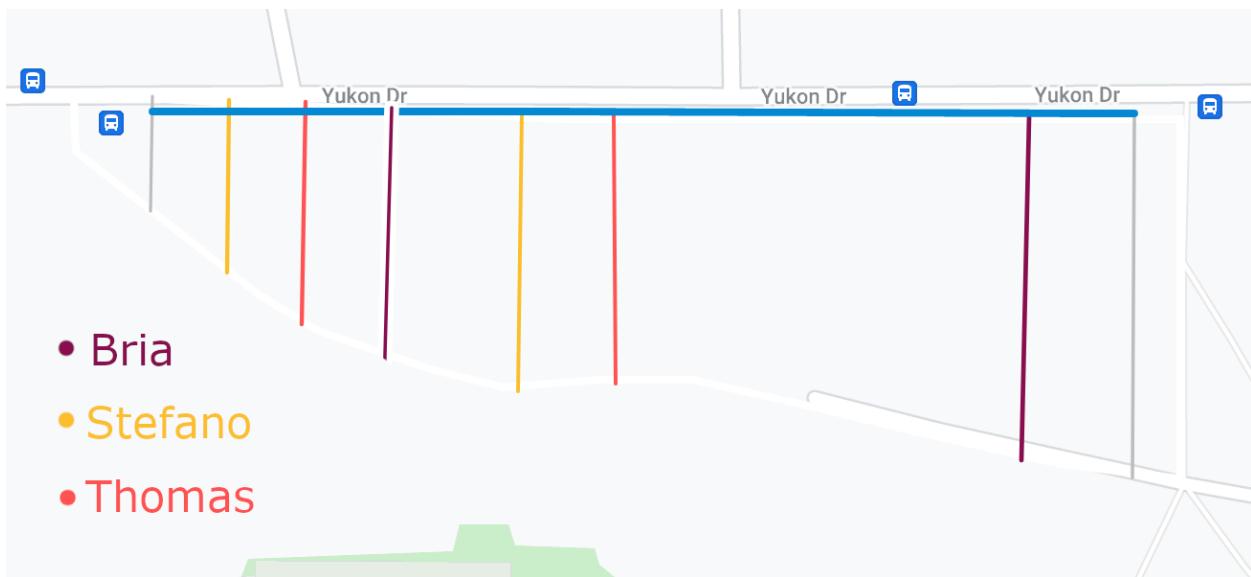
#Computing the location of samples
samp_assigned * 5 - 2.5

# samp1 samp2
#Stefano 532.5 787.5
#Thomas 722.5 452.5
#Bria 92.5 647.5

```

The midpoint of our sampled transects was computed as the length across the northernmost side of the plot starting from the west. Doing so we could create the following map to illustrate the location of our samples,

FIGURE 2. Transect Breakdown



Sampling occurred by walking down the middle of each transect and considering each tree within an arms-length. The circumference of each tree was measured using a cardboard tool in order to ensure that each tree in our sample had a circumference greater than 4 inches.



(A) Example of a Sampled Tree



(B) Example of a Not Sampled Tree

FIGURE 3. Circumference Determination

FIGURE 4. Data Table

Sampler	# of Birch	# of Aspen and Spruce
Stefano	0	10
	3	16
Bria	2	15
	0	7
Thomas	0	14
	1	5

## 2. RESULTS

From our six transects the following data was collected. Recall the formula for computing the proportion estimator with a single stage cluster sample,

$$\hat{p} = \frac{\sum_{i=1}^n \tau_i}{\sum_{i=1}^n M_i}.$$

Now recall the formula for computing the variance,

$$\hat{V}(\hat{p}) = \left( \frac{N-n}{N} \right) \left( \frac{N}{M} \right)^2 \left( \frac{MSE}{n} \right),$$

$$MSE = \frac{\sum_{i=1}^n (\tau_i - \hat{p}M_i)^2}{n-1}.$$

Recall that since we don't know M(the number of trees in the plot) we need to estimate the following term,

$$\frac{N}{M} \approx \frac{n}{\sum M_i}.$$

Computing the the estimator, variance and 95 percent confidence interval in r, we get the following.

FIGURE 5. Estimator Computation Code:

```

#Reading data
tau_i <- c(0, 3, 2, 0, 0, 1)
not_birch <- c(10, 16, 15, 7, 14, 5)
M_i <- tau_i + not_birch

#Computing Estimate
p_est = sum(tau_i)/sum(M_i)

#p_est
#[1] 0.08219178

#Computing SE
expected = p_est*M_i
mse = sum((tau_i - expected)^2)/(n-1)
se = sqrt(((N-n)/N)* (n/sum(M_i))^2 * (mse/n))

#se
#[1] 0.03303254

#Computing Confidence Interval
CI <- c(p_est + 2*se, p_est - 2*se)

#CI
#[1] 0.1482569 0.0161267

```

### 3. CONCLUSION

To summarize, our estimator gave us that the proportion of birch trees in the plot is 0.08219178 with a 95 percent confidence interval between (0.1482569, 0.0161267). From my experience sampling an exploring the plot these results seem reasonable, and thankfully if we desired a tighter confidence interval we could easily go back out to the plot and sample more clusters. Designing the clusters so that they were an arms-length wide in combination with the fact that we used a simple random sample for our clusters, made our sampling scheme very versatile. As a sampler I would say that collecting data for this projects was very straightforward, and easy while also delivering decent results for relatively small effort.