*Statistical Analysis for Silvestrum*

**Objective:**

Create a statistical analysis to find the appropriate number of sample plots in order to get an approximate calculation of total carbon per tree in Mg per hectare.

The ideal scenario would be to find the number of plots and number of trees to find a statistical analysis.

**Background information:**

Forest mangrove sizes does not seem to control how densely carbon (C) is stored in these mangroves, sampling at different locations within each forest revealed more about the distribution of C. Fringe areas had relatively low and consistent C densities, while in the interiors of forests we observed some of the greatest C densities. <https://datamares.org/stories/gulf-mangroves-store-carbon-by-the-ton/>

**Data:**

* Census data
* CSV
* Carbon calculations to Mg/ha
* Manual measurements (Canopy, Height, Crown Diameter).

|  |  |  |  |
| --- | --- | --- | --- |
| Site | Year | Number of Plots | Tree Measurement |
| A | 2010 | 7 | 746 |
| A | 2017 | 18 | 3467 |
| A | 2019 | 8 | 998 |
| A | 2020 | 46 | 4956 |
| B | 2015 | 7 | 338 |
| B | 2016 | 13 | 629 |
| B | 2017 | 11 | 439 |
| B | 2018 | 7 | 207 |
| B | 2019 | 7 | 211 |
| B | 2020 | 5 | 119 |
| Total | **10** | **129** | **12,110** |

**Summary and insights from the A/R Methodological Tool**

The tool is based in the targeted precision for the biomass stock estimate, the variability of the biomass stock in different strata within the projects area, and the number of possible sample plots within the project boundary. The number of sample plots is calculated using a formula that considers the estimated standard deviation of the biomass stock in each stratum, the relative weight of each stratum, and the acceptable margin of error. The tool assumes that the approximate area of each stratum and the approximate variance of the biomass stock in each stratum are known.

**Steps in my end**

1. Calculate the variance of biomass stock from each stratum.
   1. Have the SD of each plot
   2. Create assumptions and results for that variance. Which plot has the largest concentrated pool of carbon? And why?
      1. The less trees in the plot the more concentrated carbon we have per hectare.
   3. Maybe create a matrix with the correlations of the variables?
2. Create a function to run the AR Methodological Tool for the sample plot calculations.
   1. The formula is iterated until the final value of the number of sample plots is obtained.

**Assumptions:**

* I’m just using the plots of 154m2
* I’m merging both data sets so we can analyze more data
* We are using the power package to calculate the power

**Questions:**

1. Should I create new plots as stratums and randomize the selection?
2. Should I randomly stratify the same number of trees and calculate the mean of carbon per tree?
3. Should I calculate the mean and SD with the number of trees instead of the SD?
4. When we look at the density of trees per plot. We can see that the density is very variable and there is a trend of reducing SD as you increase density.
5. I would like to discuss the attached table that shows the summary of all plantation years (projects) with detailed information and see what else can we understand from here.
6. Interpreting A/R Tool results. I feel the results are a little bit too short comparing them to the power calculations, but they do are close to number of plots being measured.
7. Regarding the A/R Tool in the point 13, it says that if the iteration is lower than 30 then we need to continue with a second iteration. Which I would like to review with you both first. (13. If the number of sample plots *n* calculated in the first iteration using equation 1 is less than 30, then equation 1 is applied in the second iteration using the *t*-value for degrees of freedom equal to (*n-*1). The value of *n* obtained in the second iteration is the final value of *n*. )

**Things to do next**

1. Check the power calculation column showing the number of trees to have a 90% variance for each plot.
2. Check the power calculation column showing the number of plots to have a 90% variance for each year.
3. Create better plot names in the general data frames with a for loop.
4. Create the equation function and test it once you have the table variable.
5. Get deeper in understanding the height and crown size for the total variation of each tree. Understand the data. See why the results have been variant. Because everything is calculated…. So, the variance must be within the tree size. Understand the relationships between heights and canopy.
6. Explain to Silvestrum why they are so different. Higher trees have higher carbon? How is the SD behaving with a mix of small, medium, and large.
7. What is the correlation of carbon and the variation of tree height and the canopy. Create a plot of SD and the number of trees. To see how they are stepping off.