Estimating the Tree Population of ISI Bangalore

A Statistical Sampling Project

Submitted by:

Gagan Krishna S (bsd-bg-2506)

Utkarsh Bharadwaj (bsd-bg-2521)

Saurav Kumar (bsd-bg-2517)

October 29, 2025

Contents

1 Introduction

This report details the methodology and preliminary calculations for a statistics project aimed at estimating the total number of trees within the Indian Statistical Institute (ISI), Bangalore campus.

The primary objective is to apply a **stratified random sampling** technique to derive a reliable estimate of the tree population. This approach is chosen over simple random sampling to account for the heterogeneous distribution of foliage across the campus, thereby increasing the precision of our final estimate.

The project involves digital mapping using Google Earth Pro, segmentation of the campus into distinct strata, overlaying a grid system, and physical data collection.

2 Methodology

The estimation process is designed in several phases, starting with digital mapping and stratification, followed by calculation of sampling units, and culminating in physical data collection and final estimation.

2.1 Initial Data Acquisition and Mapping

The first step involved acquiring a KML (Keyhole Markup Language) file of the ISI Bangalore campus from Google Earth Pro. Using this data, the official boundaries of the campus were marked.

The campus area was then broadly segmented into functional zones, such as academic buildings, hostels, and the primary forest cover region. The focus of this study is the main forest cover area, which is outlined in **yellow** in Figure 1.

2.2 Area Segmentation and Stratification

To improve accuracy, the primary forest cover was bifurcated into two distinct strata based on observable tree density from satellite imagery:

1. **Low-Density (LD) Region:** This area, marked in **green** on the map (Figure 1), represents sections with sparser forest cover, including clearings and smaller clusters of trees.
2. **High-Density (HD) Region:** This area consists of the remaining forest cover (the total yellow-outlined area minus the green-outlined area). This region exhibits a much denser and more continuous tree population.

This stratification ensures that both types of regions are appropriately represented in the sample, preventing under- or over-estimation.

2.3 Grid System Implementation

A 10m $\times$ 10m grid (100 sq. m) was digitally overlaid on the entire campus map, as shown in Figure 2. This grid system divides the entire study area into a finite number of discrete sampling units (grids), which form the basis of our sampling frame.

Figure 1: Map showing the total forest area (yellow) and the stratified Low-Density (LD) region (green).

Figure 2: The 10m × 10m grid system (purple) overlaid on the stratified campus map.

3 Area and Grid Calculations

The total areas for each stratum were calculated from the Google Earth Pro polygons. These areas determine the total number of sampling units (grids) available in each stratum.

3.1 Low-Density (LD) Stratum

The total area of the low-density regions was calculated by summing its four constituent parts:

- Area 1: 2914 sq. m
- Area 2: 2708 sq. m
- Area 3: 1997 sq. m
- Area 4: 2162 sq. m

Total Low-Density (LD) Area = 9761 sq. m

The number of 100 sq. m grids in the LD stratum (N_{LD}) is:

$$N_{LD} = \frac{9761 \text{ sq. m}}{100 \text{ sq. m}} = 97.61 \text{ grids}$$

3.2 High-Density (HD) Stratum

The total forest cover area (yellow boundary) was calculated by summing its five main sections:

- Section 1: 10317 sq. m
- Section 2: 28320 sq. m
- Section 3: 18536 sq. m
- Section 4: 25503 sq. m
- Section 5: 4048 sq. m

Total Forest Cover Area = 86724 sq. m

The high-density area is the remainder after subtracting the low-density area from the total forest area.

Total High-Density (HD) Area = Total Forest Area - Total LD Area

$$\text{Area}_{HD} = 86724 \text{ sq. m} - 9761 \text{ sq. m} = 76963 \text{ sq. m}$$

The number of 100 sq. m grids in the HD stratum (N_{HD}) is:

$$N_{HD} = \frac{76963 \text{ sq. m}}{100 \text{ sq. m}} = 769.63 \text{ grids}$$

3.3 Summary of Sampling Frame

The total population of grids (N) is divided as follows:

- Total Grids in Low-Density (N_{LD}) ≈ 97.61
- Total Grids in High-Density (N_{HD}) ≈ 769.63
- **Total Grids in Study Area (N_{Total})** ≈ 867.24

4 Sampling and Estimation Plan

4.1 Data Collection

The next phase involved physical data collection. A number of grids were randomly selected from each stratum (LD and HD). We then physically visited each selected 10m $\times$ 10m grid and conducted a complete and accurate count of all trees within its boundaries.

4.1.1 Definition of a Tree

To ensure counting consistency, a "tree" must be clearly defined. While various local definitions exist, such as that in the Delhi Preservation of Trees Act, 1994 (which specifies $\geq$ 1m height and $\geq$ 5cm diameter at 30cm height), this study will adopt the standard used by the **US Forest Service**.

Under this definition, a woody plant is counted as a tree if it has a **Diameter at Breast Height (DBH) of 5 inches (12.7 cm) or greater**. DBH is a standard forestry measurement taken at 4.5 feet (1.37 meters) from the ground. Any woody plant not meeting this criterion will be excluded from the count.

4.2 Estimation Formula

Let:

- $N_{LD} = 97.61$ (Total grids in LD stratum)
- $N_{HD} = 769.63$ (Total grids in HD stratum)
- $n_{LD} = 3$ (Number of grids sampled from LD stratum)
- $n_{HD} = 3$ (Number of grids sampled from HD stratum)
- $\bar{x}_{LD}$ = Mean of the number of trees counted in the sampled LD grids
- $\bar{x}_{HD}$ = Mean of the number of trees counted in the sampled HD grids

The total number of trees ($\hat{T}$) will be estimated by summing the extrapolated totals from each stratum:

$$\hat{T}_{\text{total}} = (N_{LD} \times \bar{x}_{LD}) + (N_{HD} \times \bar{x}_{HD})$$

5 Results

The final estimation was done by the physical counting of trees in the randomly selected sample grids.

5.1 Low-Density (LD) Stratum Sample

- Number of grids sampled (n_{LD}): **3**
- **Mean of number of trees counted in the grid ($\bar{x}_{LD}$): 3.67**

5.2 High-Density (HD) Stratum Sample

- Number of grids sampled (n_{HD}): **3**
- **Mean of number of trees counted in the grid ($\bar{x}_{HD}$): 5.67**

5.3 Final Estimated Total

Using the estimation formula from Section 4.2:

$$\begin{aligned}\hat{T}_{\text{total}} &= (N_{LD} \times \bar{x}_{LD}) + (N_{HD} \times \bar{x}_{HD}) \\ &= (97.61 \times 3.67) + (769.63 \times 5.67) \\ &= 358.23 + 4364.80 \\ &= 4723.03\end{aligned}$$

Total Estimated Number of Trees ($\hat{T}_{\text{total}}$) $\approx$ 4723

---

6 Conclusion

This report outlines a robust stratified random sampling methodology for estimating the tree population at ISI Bangalore. By segmenting the campus into high-density and low-density strata, we have created a sampling plan that accounts for the area's natural heterogeneity, which will lead to a more precise and representative final estimate.

The final result, based on physical data collection, provides a valuable statistical snapshot of the campus's green cover.