COMPARISON OF TWO SAMPLING METHODS FOR ESTIMATING URBAN TREE DENSITY

By Ivan André Alvarez[1], Giuliana Del Nero Velasco[2], Henrique Sundfeld Barbin[1], Ana Maria Liner Pereira Lima[3], and Hilton Thadeu Zarate do Couto[4]

Abstract. Sampling can be used as a method for urban tree inventory estimation. There are several sampling methods available, and choices for urban tree inventory methods vary according to the place to be studied and the urban tree conditions. This study compared the results of simple and stratified random sampling methods with those of a total district tree census. The simple random sampling error was 17%, and the coefficient of variation was 47%. The stratified random sampling errors varied from 19% to 60%, and their coefficients of variation ranged from 32% to 70%, depending on the stratum. The Shannon diversity index (SDI) was low in the census (3.07), as in the simple random sampling (1.27). The total number of trees and the number of trees per kilometer of sidewalk calculated by the simple random sampling were similar to those obtained by the census. Because the sampling error obtained by stratified random sampling was higher than for that obtained by simple random sampling, the use of stratified random sampling was not advantageous when compared with simple random sampling. Furthermore, the stratified random sampling procedure was more complex.

Key Words. Inventory; sampling; urban trees.

Urban forest research has focused on tree surveys to assist planning by municipalities. Although a total tree survey is a difficult undertaking, there are statistical sampling methods that describe the whole from population samples.

The use of sampling as a research tool can be applied by sampling city blocks and number of trees per lineal kilometer. By means of a sampling inventory, it is possible to draw some conclusions, such as the frequency of species, their diversity, and species' adaptation to the site (Chacalo et al. 1994).

Once the limits and characteristics of the population of street trees are known and the desired level of precision is established, the use of sampling techniques constitutes a procedure of significant efficiency for street tree evaluation. Using rectangular sample plots and number of trees per kilometer of forested sidewalk as main variables, simple random sampling procedures provided a significant efficiency for street tree evaluation, as previously demonstrated by Milano (1994). However, population heterogeneity can introduce sampling errors when the simple random sampling method is used.

Stratification can be used to increase survey precision when it is extrapolated for the total population once the characteristics of the strata are taken into account. The population can be divided in subpopulations, so that each one can be more homogeneous. This division in layers is called a stratum (Cochran 1977).

Couto (1994) showed that stratified random sampling can be very useful for urban forest surveys. The strata can be selected based on the district, street tree density, or a group of blocks, depending on the criteria chosen for grouping the sampling units for stratum composition. To obtain a coherent survey, the strata should possess a homogeneity of measured values. Thus, a precise median value of any stratum estimate can be obtained using a small sample of that stratum. According to Cochran (1977), those estimates can be combined to produce an accurate estimate of the total population.

Jaeson et al. (1992) used a method that combined different-sized blocks to identify the strata. The mean number of trees per randomly selected block was multiplied by the number of trees per block to estimate the total number of trees.

The choice of sampling type depends on a previous analysis or pre-sample of the area to be studied. According to Milano et al. (1992), it is necessary to quantify the forest and know its distribution in urban areas by defining its characteristics and quality.

Determining the purpose of the inventory is important in choosing the most appropriate methodology because each methodology presents a different degree of precision (Grey and Deneke 1978). According to Couto (1994), determining the most appropriate sampling type depends on the distribution of the measurements in space and time. A common criterion for choosing sampling techniques is species present and their relative frequency.

This study compares simple and stratified random sampling methods with a total tree survey of a district that has varying characteristics.

MATERIALS AND METHODS

The city of Piracicaba is located at 22°42′ 30.9″ S latitude and 47°38′ 01″ W longitude. The tree inventory was performed in Santa Cecília district (Figure 1), located in the East Zone of Piracicaba, São Paulo, Brazil. The residents of Santa Cecília are in the middle- and upper-middle-class income groups.
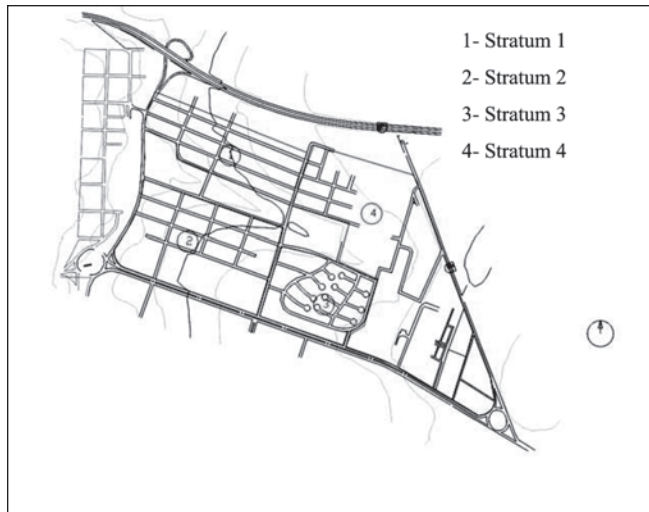
Figure 1. Map of Santa Cecília district, Piracicaba, São Paulo, Brazil.

The district of Santa Cecília contains parcels of land that were occupied gradually. The urban district is relatively new. In 1970s, urbanization began with intensity and continues. This fact has contributed to the diversity of urban characteristics with regard to human population, socioeconomic level, and presence of vegetation. This heterogeneity must be analyzed in full.

The block was the sampling unit chosen, and the data records were expressed as number of trees per lineal kilometer of sidewalk. The variable definition was the ratio of the total number of existing trees on the sidewalks to the total of kilometers of sidewalk. AutoCAD software measured these values to obtain a district map.

The total number of blocks in Santa Cecília district is 57, excluding public free spaces. An initial 10 random blocks were sampled, with the later addition of 11 more random blocks. One block was dropped from the total of 21 because it had no trees and would thereby exert undue influence on the analysis. For the stratified random sampling, the same 20 blocks were divided into four strata. The sampling strata were divided to achieve homogeneity, based on date of initial occupancy, physical proximity to other blocks in the stratification group, and purchasing power of residents.

The typical sampling unit was a four-sided block, but other clearly defined shapes were acceptable, especially in stratum 3, which had cul-de-sacs.

The measured variables were number of trees per lineal kilometer of sidewalk and total number of trees in the district. The estimate of the total number of trees in the district (57 blocks) is important, even though it does not express a difference in the tree density. The number of trees per kilometer of sidewalk gives a clearer measure of the presence of the trees per human-occupied space. On the other hand, evaluating the density does not mean that the quality of the urban forest is analyzed (i.e., there could be an excess of nonappropriate species and/or low diversity). Therefore, the Shannon diversity index (SDI) was selected to describe species diversity in relation to the total number of trees (Pielou 1975).

Simple Random Sampling

The number of trees per kilometer of sidewalk was estimated according to Cochran (1977), who defines population ratio as

$$R = \frac{X_T}{Y_T},$$

where X_T is the number of existing trees in the blocks, and Y_T is the number of lineal kilometers of sidewalk in the blocks.

The sampling ratio is represented by

$$r = \frac{\sum_{i=1}^{n} x_i}{\sum_{i=1}^{n} y_i},$$

where x_i represents number of trees in the selected block, and y_i represents the total number of kilometers of the selected block. The interval of 95% of trust for each one of the population ratios is given by $[r - 2s(r), r + 2s(r)]$.

The variation coefficient is represented by

$$\frac{\sqrt{s^2}}{\overline{x}},$$

where $\overline{x}$ is the sample average, and s^2 is the variance, whose formula is

$$s^2 = \frac{1-f}{n\overline{y}^2(n-1)}\left(\sum_{i=1}^{n} x_i^2 - 2r\sum x_i y_i + r^2 \sum_{i=1}^{n} y_i^2\right).$$

The sampling fraction is represented by

$$f = \frac{n}{N},$$

where N is the total number of blocks in the studied area, and n is the number of blocks selected for the sample.

The median value of the total number of kilometers variable is represented by

$$\overline{y} = \frac{\sum_{i=1}^{n} y_i}{n},$$

and the sampling error, expressed as a percentage, is represented by

$$\text{Sampling error (\%)} = \frac{t \cdot s\overline{x} \cdot 100}{\overline{x}}$$

Stratified Random Sampling

According to Cochran (1977), the number of trees per kilometer in stratified random sampling is estimated by the separate and proportional population ratio. The sampling ratio is the same in all strata: denominated stratification with proportional partition of the blocks per stratum.

The formula of this ratio is

$$\hat{R}_{es} = \frac{\left(\sum_{h=1}^{H} \frac{x_h}{y_h} Y_h \right)}{Y_t}.$$

The components of the preceding formula are

$$x_h = \sum_{i=1}^{n_h} x_{ih} \text{ and } y_h = \sum_{i=1}^{n_h} x_{ih},$$

where x_{ih} is the number of existing trees in the i block of the h stratum, and y_{ih} is the value of the lineal kilometers of sidewalk of the h stratum. The total number of kilometers of sidewalk of the h stratum is

$$Y_h = \sum_{j=1}^{N_h} Y_{jh},$$

where N_h is the total of blocks of the h stratum, and Y_{jh} is the total of kilometers of sidewalk of the j block of the h stratum. The total population of the kilometers of sidewalk variable is

$$Y_T = \sum_{h=1}^{H} Y_h$$

The coefficient of variation, the sampling ratio per stratum, and the sampling error for each one of the population ratios per stratum were obtained as in the simple random sampling. The Shannon diversity index (SDI) was calculated, describing the variability of the species number for the total number of trees. The SDI is calculated by the following formula:

$$H = -\sum_{i=1}^{S} p_i \ln(p_i),$$

where p_i is the proportion of individuals found in the i species, and ln is the natural logarithm of p_i.

RESULTS AND DISCUSSION

Table 1 shows the census, the simple random sampling, and the stratified random sampling results.

The total number of trees and the number of trees per kilometer of sidewalk obtained by simple random sampling are similar to the real numbers obtained by the census. The stratified random sampling values, considering the average of all strata in the total area and obtained from the strata of the total number of trees, presented a larger error in relation to the census results. It is important to remember that the total number of trees was not simply calculated as the sum of the individual values of each stratum but was calculated as the strata average multiplied by the number of strata.

The evaluation of the number of trees should be described together with the analysis of the sampling error. The simple random sampling error is 17%, which was not considered high. The error calculation is important to give an idea of the usefulness of the sampling method for an urban forest inventory. The sampling error values of each stratum were indicators that not all the strata were homogeneous. The sampling error that represented the stratum homogeneity varied from 19% (stratum 1) to 60% (stratum 4). Where sampling error was lower, the number of trees per kilometer of sidewalk was closer to the real value. In strata 1 and 2, where the tree/km of sidewalk average was a little higher than in strata 3 and 4, the sampling errors were also higher. Stratum 1 presented a higher trees per kilometer of sidewalk value (36.22) than stratum 4, which had the lowest of all (15.16). The strata general average was not similar to the census value; however, stratum 1 results approached it.

In the stratified sampling, the estimated number of trees per kilometer of sidewalk (779 trees) represents the total number of trees of the 57 blocks, while the average number of trees per kilometer value represents the number of trees per kilometer of sidewalk of the 20-block sample.

The coefficient of variation represents the precision of the experiment. When the purpose is population sampling, it depicts the sampling validity. The simple random sampling presented a high coefficient of variation (47%), demonstrating an imprecision with regard to sample choice. That variability was due to the variation of the samples relative to each other, but evaluating the entire sampling results provided an accurate estimate when compared to the real values. This is what Cochran (1977) called imprecise and exact sampling.

Table 1. Inventory data from the three methods applied in the district of Santa Cecilia, São Paulo, Brazil.

| | Census | Simple random sampling | Stratified random sampling | | | | |
			Total	Stratum 1	Stratum 2	Stratum 3	Stratum 4
Total number of trees	958	900	779	297	217	141	124
Trees per lineal kilometer	37.66	35.38	30.64	36.22	26.48	17.18	15.16
Coefficient of variation (%)		46.77		32.30	51.11	37.20	70.52
Sampling error (%)		17.32		19.35	29.51	38.98	60.39

When the strata were analyzed, a higher precision was verified in stratum 1, and stratum 4 showed the highest imprecision. This finding demonstrates that stratum 1 possessed the highest sampling homogeneity, which was confirmed in the field.

The difficulty of obtaining a homogeneous stratum to maintain the criteria previously defined for all strata was shown only in stratum 4. The lots in stratum 4 were found to have unoccupied areas that did not possess any urban forest. Areas with new construction presented a high density of trees, elevating the number of trees in the stratum.

Table 2. Species in the census and in the simple and stratified random samplings.

Species	Census	Simple random sampling	Stratified random sampling			
			Total	Stratum 1	Stratum 2	Stratum 3
Murraya paniculata	15.97	19.75	36.30	0.94	12.00	13.16
Caesalpinia peltophoroides	12.63	14.33	15.75	13.21	20.00	7.89
Lagerstroemia indica	8.98	7.01	2.05	8.49	8.00	21.05
Tabebuia spp.	8.98	8.60	3.42	10.38	12.00	21.05
Schinus molle	6.68	2.23	2.05	3.77	—	—
Terminalia catappa	4.18	3.82	4.11	5.66	—	—
Thevetia peruviana	3.86	3.50	4.11	2.83	—	5.26
Licania tomentosa	3.65	3.82	2.05	8.49	—	—
Tibouchina granulosa	3.44	2.55	4.11	0.94	—	2.63
Syagrus romanzoffiana	3.24	5.10	—	10.38	—	—
Nerium oleander	3.13	1.91	1.37	3.77	—	—
Hibiscus rosa-sinensis	2.71	3.50	0.68	9.43	—	—
Delonix regia	2.51	—	—	—	—	0.00
Ficus microcarpa	1.88	4.14	6.85	1.89	—	—
Enterolobium contortisiliquum	1.77	—	—	—	—	—
Others	16.39	19.75	17.12	19.81	48.00	26.32

The comparison of a simple and a stratified random sampling on the same blocks raised important facts that could be analyzed purely by their quantitative aspects, in a statistical way, as well as by the qualitative aspects of the urban forest. The identification of the species used in the urban forest was important to define the choice of future species that could be planted. Table 2 shows the species selected in the sampling schemes compared to those verified by the census. Not all the found species are listed—only the main ones obtained in the census.

The Shannon diversity index gives an idea of how much species diversity is enough when compared with the total number of trees, as shown in Figure 2.

An analysis of Table 2 demonstrates that a few species occupy most of the urban forest. The census found that 15 species comprise 84% of the number of trees, while in the simple random sampling and strata 1, 2, 3 and 4, the same species comprised 80%, 83%, 80%, 52%, and 74% of the trees, respectively. Tree species were not evenly distributed within the total district; thus, not all species in the census were found in each sampling unit. It is important to point out that the diversity in stratum 3 was much lower than the others. This fact is probably related to the lower economic level and the poor quality of tree planting in that area.

The Shannon diversity index was a little higher in the census than in the simple random sampling and was much lower in the stratified random sampling (Figure 2). On the other hand, the simple random sampling presented an SDI lower than 3. According to Martins and Santos (2001), the index uses sampling that includes all the species, but this ideal situation is impossible once there is a high degree of heterogeneity in urban forest. This fact can explain the differences between the samplings and the census values.

A street tree population is complex, with many urban factors interacting to influence the frequency and diversity of that population, thus making it difficult to choose an appropriate inventory method for all situations. Each case is unique, but, in several studies in Brazil, such as that of Rachid and Couto (1999), the stratified random sampling was no better than simple random sampling. In our case, simple random sampling was an advantageous technique because stratified random sampling is a more complex method.
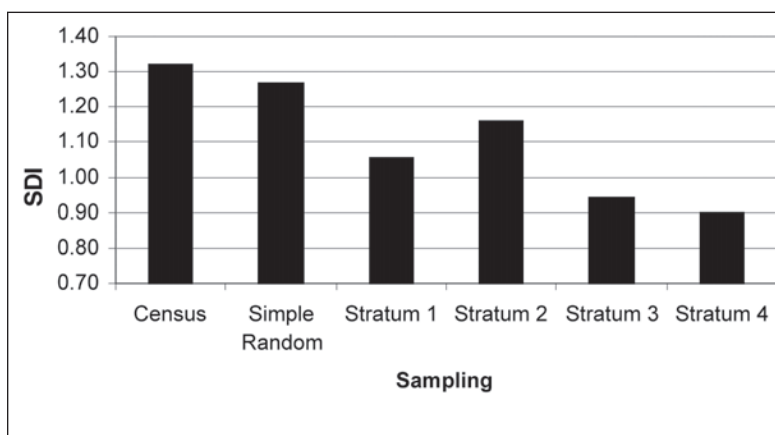

Figure 2. The Shannon diversity index (SDI) for the census and for the simple and stratified random samplings.

CONCLUSIONS

The simple random sampling of the district was more appropriate to estimate the total number of trees and the number of trees per kilometer of sidewalk than the stratified random sample. However, the large coefficient of variation showed that the sampling distribution along the district was not uniform. We concluded that systematized sampling would be more appropriate than simple random sampling.

The stratified random sampling did not represent the sampled universe in a reliable way. There is a need to establish other criteria for strata block grouping that differ from those used in this work. Stratification may not be a suitable method for urban forest analysis because of the multiplicity of variables and randomized factors that determine the presence of trees in a city that such sampling cannot express.

LITERATURE CITED

Chacalo, A., A. Aldama, and J. Grabinsky. 1994. Street tree inventory in México City. J. Arboric. 20(4):222–226.

Cochran, W.G. 1977. Sampling Techniques (3rd ed.). Wiley, New York, NY. 428 pp.

Couto, H.T.Z. 1994. Métodos de amostragem para avaliação de árvores de ruas. pp. 169–179. In Congresso Brasileiro de Arborização Urbana 2, Encontro Nacional Sobre Arborização Urbana 5, São Luis. Sociedade Brasileira de Arborização Urbana, São Luís, Brasil.

Grey, G.W., and F.J. Deneke. 1978. Urban Forestry. Wiley, New York, NY. 279 pp.

Jaeson, R., N. Bassuk, S., Schwager, and D. Headley. 1992. A statistical method for the accurate and rapid sampling of urban street tree populations. J. Arboric. 18(4):171–183.

Milano, M.S. 1994. Métodos de amostragem para avaliação de arborização de ruas, pp. 163–168. In Congresso Brasileiro de Arborização Urbana 2, Encontro Nacional Sobre Arborização Urbana 5, São Luis. Sociedade Brasileira de Arborização Urbana, São Luís, Brasil.

Milano, M.S., O. Sarnowiski Filho, and J.A.M. Robaio. 1992. Estudo comparativo de unidades amostrais utilizadas para inventário quali-quantitativo da arborização de rua em Curitiba, pp. 343–350. In Encontro Nacional Sobre Arborização Urbana, 4, Vitória. Sociedade Brasileira Urbana, Vitória, Brasil.

Pielou, E.C. 1975 Ecological Diversity. Wiley, New York, NY.

[1*]*Ph.D. in Plant Production*
Travessa Francisco Gobeth, 65
Piracicaba – SP/Brazil
CEP: 13.418-030

[2]*Master of Science in Plant Production, USP – ESALQ*
[3]*Professor of Department of Plant Production*
[4]*Professor of Department of Forest Resources*
Universidade de São Paulo (USP)
Escola Superior de Agricultura Luiz de Queiroz (ESALQ)
C.P. 9, CEP: 13418-900 – Piracicaba, SP

*Corresponding author: Ivan André Alvarez (iaalvare@ terra.com.br).

Résumé. Le choix entre chaque type d'échantillonage, comme l'aléatoire simple et l'aléatoire en couches, par exemple, varie en fonction du lieu et des conditions de l'arborisation urbaine. Cette étude a comparé les résultats des deux méthodes d'échantillonage avec l'étude complète de tous les arbres du district (recensement). L' erreur de l'échantillonage aléatoire simple était de 17% et le coefficient de variation était 47%. Les erreurs de l'échantillonage en couches ont varié entre 19% et 60% et ses coefficients de variation entre 32% et 70%, selon la strate. L'Index de Dispertion de Shannon (IDS) était bas dans le recensement (3,07) comme dans l'échantillonage simple (1,27). Les résultats du nombre total d'arbres et du nombre d'arbres par kilomètre de trottoir calculés par l'application de l'échantillonage simple sont proches de ceux obtenus par le recensement, et ce type d'échantillonage a été considéré adéquat . Par contre, les résultats obtenus par l'application de l'échantillonage en couches montraient une grande différence par rapport à ceux du recensement. En conclusion, les résultats de l'échantillonage en couches ne sont pas fiables pour représenter l'univers étudié.

Zusammenfassung. Die Auswahl von einer Stichprobe für eizelne Experimente, wie zum Beispiel, eine zufällige Stichprobe, variert in Zustimmung mit dem Ort und Klima wo die Bäume wachsen. In dieser Arbeit verglich man die Ergebnisse von einigen Stichproben in Bäume die in einen Stadtteil wachsen. Der zufällige Fehler der Stichproben war 17% und der variation-koeffizient war 47%. Die geschichtete Fehler der Stichproben variierten zwischen 19% bis 60%, und Ihre Variation-Koeffizienten variierten von 32% bis 70%, und hingen von der Schichtung ab. Das Shannon Diversifizierende Register (SDR) war niedrig in der Zählung (3,07) und in der zufälligen Stichproben (1,27). Die Ergebnisse der Anzahl von Bäumen in den Strassen war ungefähr gleich in der Zählung und in der zufällinge Stichproben, so dass solche Schätzungen richting waren. Grössere Unterschiede wurden festgestellt als man Vergleichungen zwischen den Schichstproben und den Zählugen studierte, so dass solche Schätzungen nicht zuverlässig waren.

Resumen. La opción del tipo de muestreo utilizado para cada caso, como por ejemplo, el muestreo aleatorio y el estratificado, varía de acuerdo con el lugar a ser estudiado y las condiciones en que se encuentran los árboles urbanos. Este estudio ha comparado los diferentes tipos de muestreo con el estudio total (censo) de los árboles en un barrio. El error en el muestreo aleatorio ha sido de 17% y el coeficiente de variación 47%. Sin embargo, en el estratificado el error ha variado de 12% a 21,5% y su coeficiente de variación ha oscilado de 25% a 95%, dependiendo del estrato. El Índice de diversidad del Shannon (IDS) ha sido bajo tanto en el censo (1,32), como en los muestreos aleatorios (1,27) y en los estratos 1, 2, 3 y 4, respectivamente, 1,06, 1,16, 0,94, 0,90. Los resultados referentes al número total de árboles y el número de árboles por kilómetro de acera encontrados en el muestreo aleatorio se han acercado a los obtenidos por el censo, demostrando que este tipo de muestreo es considerado apropiado para tal valoración. No obstante, los resultados obtenidos por la aplicación del muestreo estratificado, han tenido una diferencia mayor con respecto a los adquiridos por el censo, así pues, el muestreo estratificado no es fiable para representar el universo probado.