AREA SAMPLING ILLUSTRATED: MEASURING TREES IN A WOODED AREA

C. H. Proctor, North Carolina State University
Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203

Background

Area sampling was developed in the 1940's, as described by A. J. King (1945), to collect information on U.S. agriculture and shortly became adapted by the Bureau of the Census for collecting household information. Its resounding success, down through the years in the Agricultural Enumerative Surveys and the Current Population Surveys, may suggest that its usefulness depends on the scale of the survey being sufficiently large and being repeatedly administered to justify the expense of acquiring detailed maps and size information. The point of this case study is to illustrate that frame construction costs can be kept low if enumerator costs are allowed to rise and area sampling can be efficient even for moderate sized surveys and particularly with a time constraint.

The methodology for constructing count units, acquiring size information and then allocating sampling units (SU's) to count units is well described in Monroe and Finkner's (1959) *Handbook of Area Sampling* and will be illustrated below as well. The necessity for blindly and rigidly following the steps for doing the area sampling arithmetic can hardly be overemphasized. These include (1) cumulating the sizes by count units, then (2) settling on the total number of sampling units N as a conveniently rounded integer giving a compact cluster size somewhere near optimum (see Proctor (1992) *Basic Cluster Sample Designs*), (3) multiplying cumulative sizes by the ratio of N over total size to get cumulative SU's, (4) randomly selecting SU numbers for the sample and finally (5) translating the selected numbers into instructions to enumerators (or to one's self as in this case) for locating land areas for segmenting and rerandomized selection of an SU to be observed.

These five clerical steps are sandwiched between creation of the count units on map materials and the field work, both of which ordinarily require considerable ingenuity, but relatively little expenditure of time. Combining this rough-and-ready style of field work with a sample design in replicated subsamples (Deming (1960) *Sample Design in Business Research*) for which the estimates and their standard errors are easily computed, renders the methods ideal for an enumerative survey with a focused objective. In its strict use of randomization and the advantages of hidden stratification from the geographic arrangement of the count units it may even serve for analytical surveys as well.

Since the advantages of simplicity are lost if one uses unequal selection probabilities for the count units or does systematic subsampling from lists within count units, these design features should be recognized as quite different. That is, the method we are describing is a one-stage design. There are oftentimes good reasons for multi-stage designs but we would suggest that one always consider this compact cluster ("ultimate" cluster) one-stage design as the default proposal.

Terminology Review

As with any scientific sampling method we must first construct a frame. A distinctive feature of area sampling is the nested structure of the frame and its embodiment in maps. The sampling unit (SU) is also called an area segment. It consists of a well demarked compact piece of land. Area segments are nested in count units, the next larger sized piece of land, while area segments contain the elements. Elements may be even smaller pieces of land ("points" even), or they may be the objects or organisms, on which measurements are made.

It is necessary that the boundary of the total area to be surveyed be well defined and drawn on the map. It is also necessary that the boundaries of the count units be well defined and drawn on the map. However, the area segment boundaries need not appear until later. It is necessary that some guess of the number of elements in each count unit be provided. These may be furnished by some expert, by a previous census, or can be found by cruising as is done for this illustration, or by any reasonable method.

Constructing Count Units

In the present case the elements are trees and the area to be covered is a woods around two dormitories. The boundaries are a creek and two roads. One can note, by cruising the area, some trails, another smaller creek, rocks and other landmarks inside the area to serve as count unit boundaries. If the boundaries of the area to be surveyed can be found on a map or on an aerial or satellite photo, then sketching them in large scale on a sheet of paper is a good thing to do before cruising. I thought I knew the outlines and sketched them without referring to a campus map. They are as shown in Figure 1.

I walked down Morrill Drive to the creek and along a trail parallel to the creek to Western Blvd. which I then followed back to Morrell Drive. I counted trees (the 40, the 120, the 160 in the upper left corner of Figure 1) until I had a rough notion of how many would be found in various sized areas. The remaining tree counts were entered on to Figure 1 by judgement and verified by a few actual counts.

Figure 1 was next redrawn with more careful attention to the actual outline of the area and with count unit (CU) boundaries. The list of count units with their estimated sizes is also shown in Figure 2. We can note that the first total was 1750 (crossed out) so the last 50 for CU 17 was changed to 100 to make the total a more "rounded" 1800. This should serve to illustrate the status of the "judged sizes" as approximations. Insofar as the 50 is more correct than the 100, sampling units in CU 17 will be a little smaller than in other places, but since there will be more of them, no bias will result.

The next step of cumulating the judged sizes is part of the obligatory area sampling arithmetic. The final step of assigning numbers of sampling units to count units requires that we specify the size of a sampling unit in terms of the judged size numbers. In class discussion we had balanced guessed costs against guessed adjacency correlation and arrived at a size of 6 trees for the area segment. This implies the whole wooded area is to be divided into $N = 300$ area segments.

Sample Selection and Enumeration

It may not be evident, but at this stage we have assigned a given number of SU's to each CU. Dividing cumulative sizes by 6 and rounding produces cumulative SU's. For example, dividing 40 by 6 and rounding yields 7 SU in count unit No. 1.

We used coin flips to select a number at random in the range 1 to 600 and let's illustrate the method. First find the power of 2 needed to exceed N. For $N = 600$ this is $2^{10} = 1024$. Then flip a coin 10 times and record the sequence of the H's and T's to get the random "word." Translate the word into zeroes (H's) and ones (T's) and convert it, as a binary number, to decimal and add one.

We flipped the word HHTTHTHHHH, which becomes 0011010000 and equals $0 \times 512 + 0 \times 256 + 1 \times 128 + 1 \times 64 + 0 \times 32 + 1 \times 16 + 0\text{'s} = 208$. The final random number is thus $208 + 1 = 209$. Note that the cumulative SU's for count units 10 and 11 are 198 and 215 so SU No. 209 is in count unit No. 12. We refer to the map to see where is CU 12. It looks to be pie shaped with interior corners at some rocks and at a tree with a ribbon. There are $215 - 198 = 17$ SU's in it.

At this point, with Figure 3, we still have a fairly larger number (17) of area segments to define and so one may choose to subdivide the count unit into parts. A rough rule is to make the number of parts equal to the square root of the number of area segments. Thus we used 4 parts as shown on Figure 3.

The boundaries in Figure 3, other than the curving trail, are lines of sight between the landmarks noted as "big tree," "cedar," "pile" (of wood), etc. The 17 area segments are assigned as shown (3, 4, 4 and 6) to the parts. A random word was generated as HHTTH which becomes $00110 = 4 + 2 = 6$ and yields $6 + 1 = 7$.

Figure 4 shows the selected part which had a prominent rock near its center and allowed a convenient division into four quadrants as the area segments. We labeled the quadrants as HH, HT, TH and TT and by coin flip got TH. Figure 5 is a bit of a jumble, but it shows the locations of the 11 trees in the selected area segment. The vertices of the triangular SU are the center rock, a tree with 2 red ribbons (Plate I), and a plastic cup on the trail (Plate II).

We also show the height and diameter measurements on the trees as well as the board feet and cords derived from them. Finally these are converted to a "price" of $170. Is it clear that if one were to repeat this operation 599 times for the other SU's then the sum of 600 "prices" would equal the total price of the woods? Indeed it would.

Plate I. The tree with 2 red ribbons

Place II. The plastic cup.

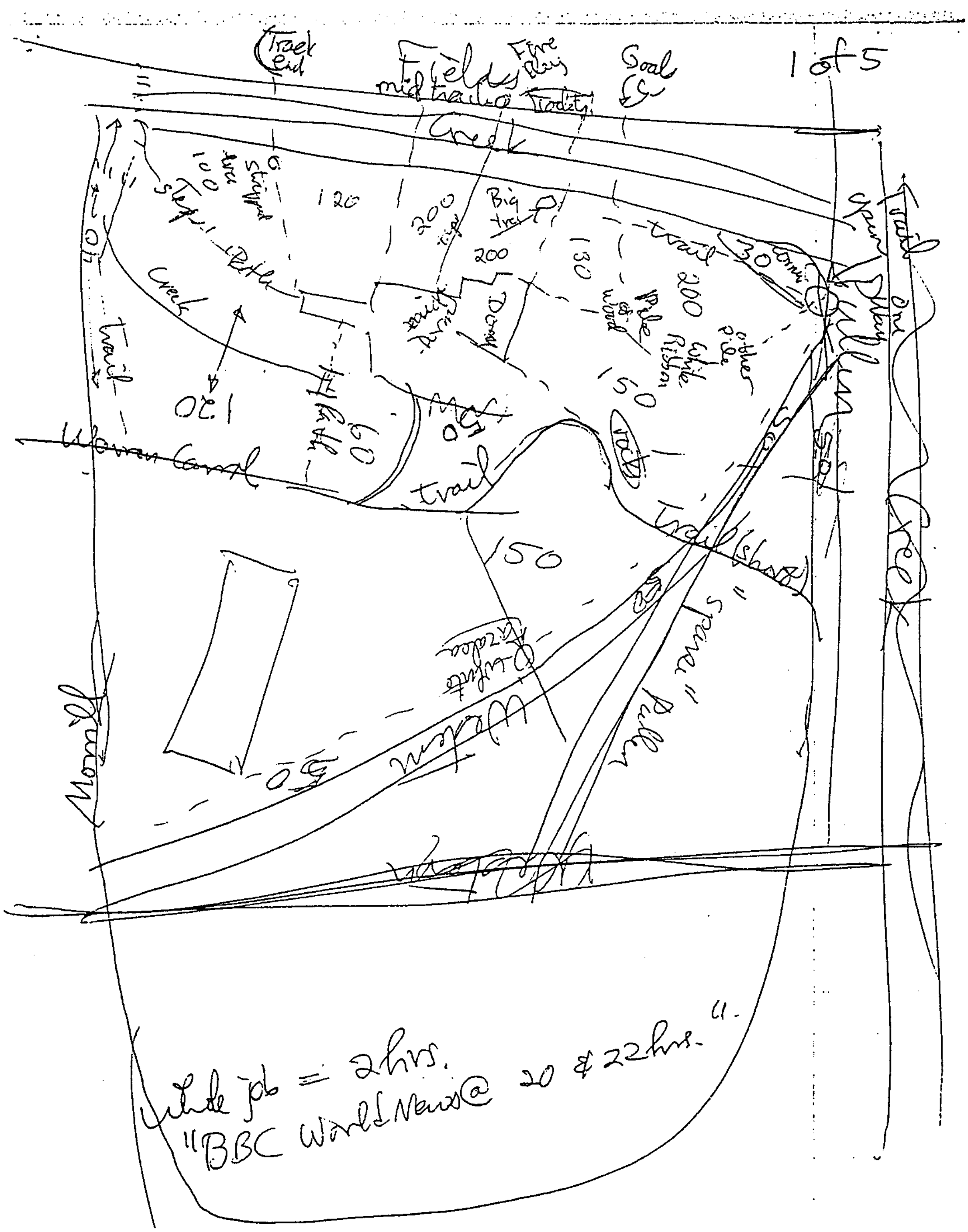

Figure 1. Preliminary sketched outline showing tree counts from cruise.

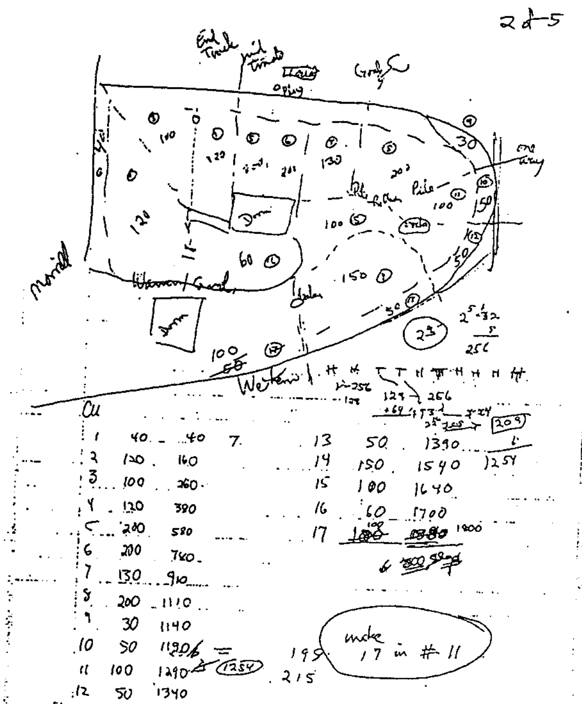

Figure 2. Revised outline with area sampling arithmetic and coin flips for selection.

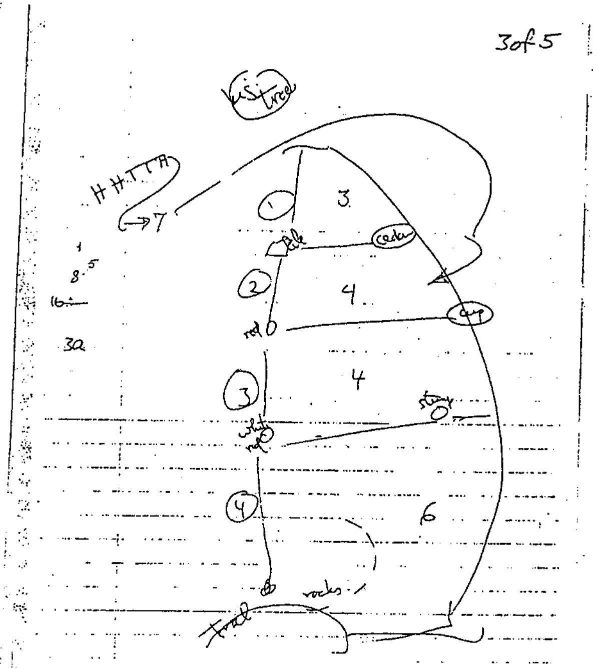

Figure 3. Selected count unit, with 4 parts defined and SU's assigned to parts. Coin flips select second part.

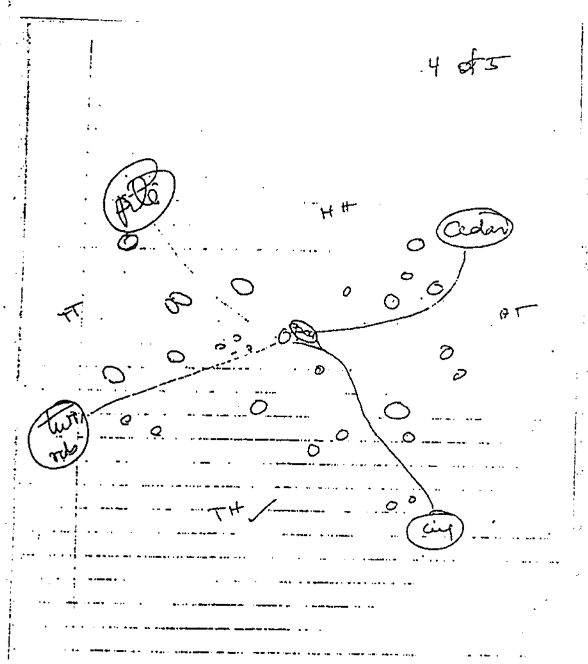

Figure 4. Subdivision of second part into area segments and selection, by 2 coin flips, of TH.

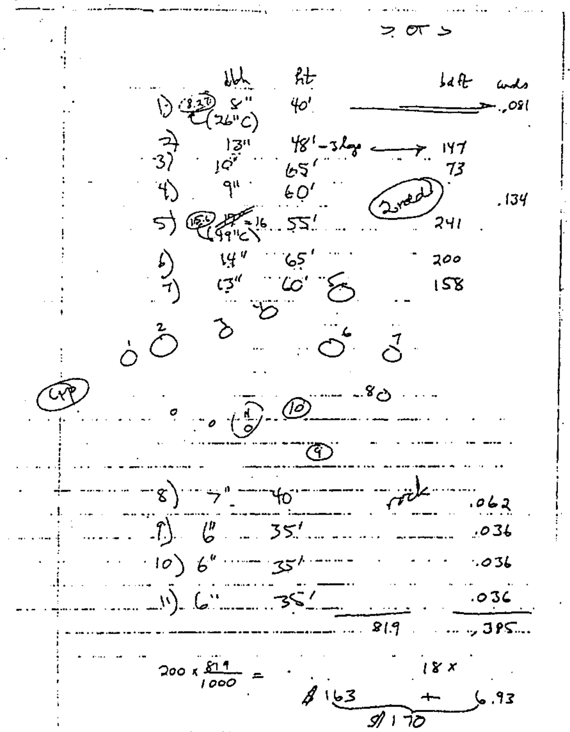

Figure 5. Tree locations in selected sampling unit with measurements and pricing of the salable trees.