NURSERY INVENTORY WORKSHOP

R. P. Karrfalt and O. Hall[1/]

Abstract.--This paper covers the contents of a one hour workshop presented at the nursery conference. The topics of the workshop were: graphic and statistical description of variability, confidence to be placed in the accuracy of inventory estimates, application procedures for systematic and random sampling, history plots, and, controlling nursery bed variation with management practices.

Additional keywords: Variation, history plot, nursery management, numbers of samples, random sampling, systematic sampling.

How many trees do we have in the nursery? This is an easy question to ask, but often a hard one to answer. To obtain the answer, a careful inventory needs to be conducted.

The most accurate way to inventory is to count all the trees. It goes without saying that this cannot be done because it is too time consuming. Therefore, we will count only some of the trees, or, in other words, we say we will count only samples of the trees.

There are many shapes our samples might have, such as circles, squares, or single rows of trees. A sample shape that is easy to use and avoids some theoretical problems is the 1 x 4 foot sample.

We must first discuss some basic statistical concepts which are very necessary to use if we are to understand our counts of seedlings. With these basic concepts, we can discuss the application of three types of inventory: systematic plots, random plots, and history plots. We will conclude our workshop by discussing the relationship between management practices and inventory data.

VARIABILITY

Variety might be the spice of life, but variability is the hard part about nursery inventory. However, it is from our understanding of variability that we will be able to understand the merits of the different sampling procedures and be able to conduct accurate inventories while keeping costs as low as possible.

In an ideal world, the nursery bed would have only plantable seedlings growing in it, and there would be the same desired number of seedlings per square foot. The nurseryman could plant 31 seeds, evenly spaced on each square foot of bed, and all 31 would germinate and give 31 plantable seedlings. In such a world, inventory might not even be necessary. But if we did do one, we

[1/] Seed Processing Specialist and Seed Testing Specialist, National Tree Seed Laboratory, Region 8, USDA Forest Service, Dry Branch, GA

would only need to measure one square foot and then to multiply the number of seedlings counted by the number of square feet of bed. An ideal nursery bed would be diagramed as in figure 1. There is no variability in our ideal nursery.

31	31	31	31
31	31	31	31
31	31	31	31
31	31	31	31

Figure 1.--A diagram of an ideal nursery bed with no variability.

In real life, variability is everywhere. The normal distribution is often a useful and appropriate way of describing the variability found in biological systems. To understand the meaning of distribution, we will look at some simple examples where we will draw some bar graphs.

Figure 2 shows a diagram of an extremely uniform nursery bed. Each linear foot is marked in this diagram, and the number of seedlings per square foot is shown on each linear division. (The number of seedlings per bed foot can be used in place of the number of seedlings per square foot. However, in this workshop we will use the per square foot term.) In our diagram there are 20 plots with 28 seedlings per square foot (spsf), 9 plots with 27 spsf and 11 plots with spsf. A bar graph of these counts, or frequencies, is shown in figure 3.

28	28	27	31	27	28	28	28
27	28	31	28	28	31	28	27
28	31	28	27	31	31	28	31
31	31	28	27	28	27	27	28
31	27	28	28	28	31	28	28

Figure 2.--Diagram of a hypothetical nursery bed showing the number of seedlings per square foot for each linear foot of bed.

Figure 4 is a diagram of a 100 foot nursery bed like the diagram in figure 2. We can make a bar graph (figure 5) from these counts. By drawing a smooth line across the top of the bars in figure 5 we have an approximate shape of the normal distribution and its relative, the t distribution.

There are some useful calculations that can be made for the normal distribution that will guide us in determining the precision of our estimates of numbers of trees and also on how many sample plots we should take.

The first calculation is for determining the sample mean $\overline{X}$. This is commonly called the average.

$$\overline{X} = \frac{\Sigma X}{n}$$

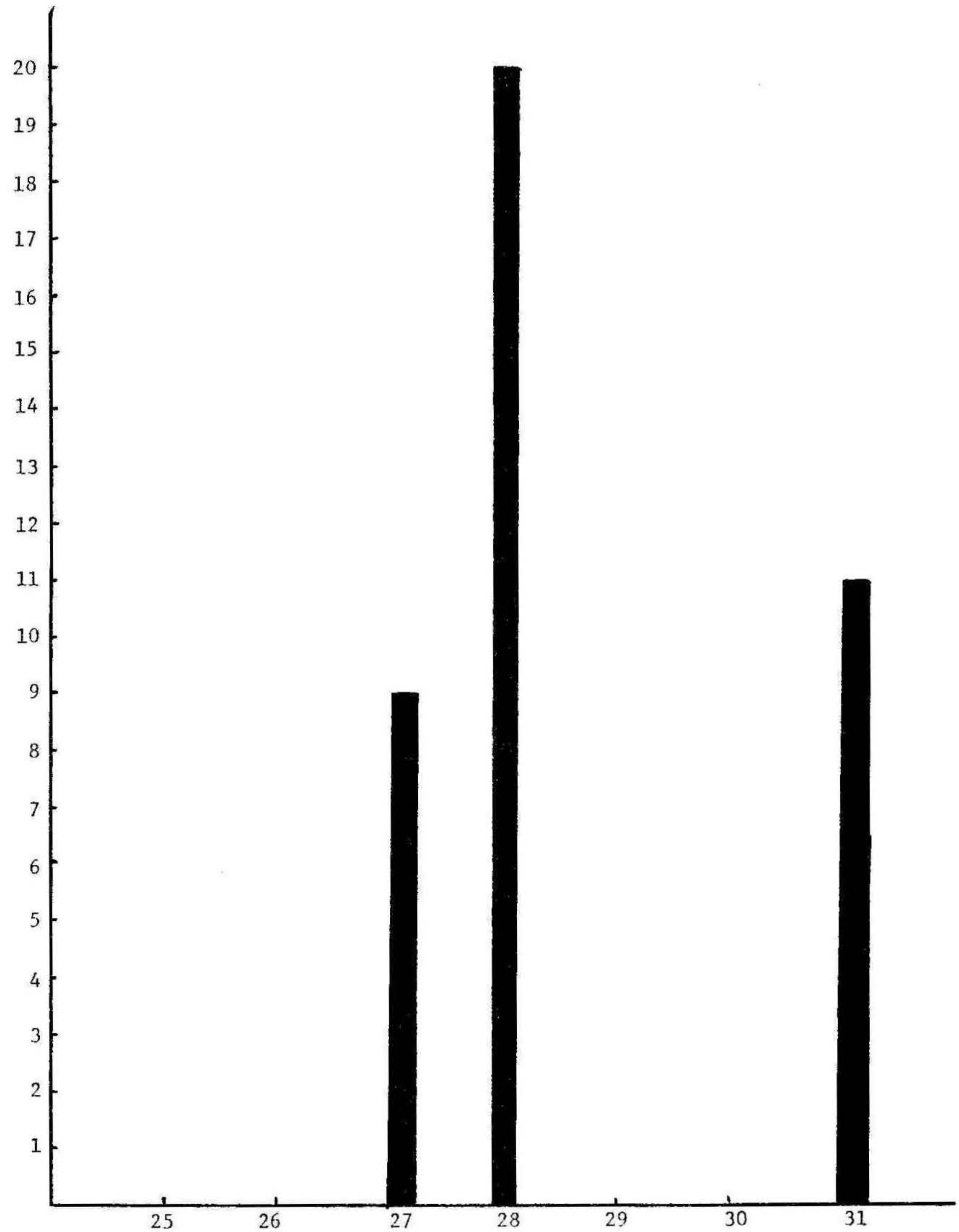

Figure 3.--Bar graph of frequencies of seedling densities in figure 2.

30	32	38	28	30
25	31	36	30	29
38	30	32	29	40
40	34	29	31	31
24	33	28	32	32
18	33	27	33	30
20	32	30	32	29
27	30	45	30	27
31	32	31	29	26
32	34	29	30	28
29	31	30	28	31
37	29	26	29	33
39	28	27	27	29
40	30	30	30	27
29	29	31	29	28
32	32	34	28	29
35	33	35	28	29
31	31	32	28	30
29	30	31	27	31
24	29	28	29	30

Figure 4.--A diagram of a hypothetical nursery bed, 100 feet long, with low variability, showing the number of seedlings per square foot for each linear foot. Called bed 1 in the text.

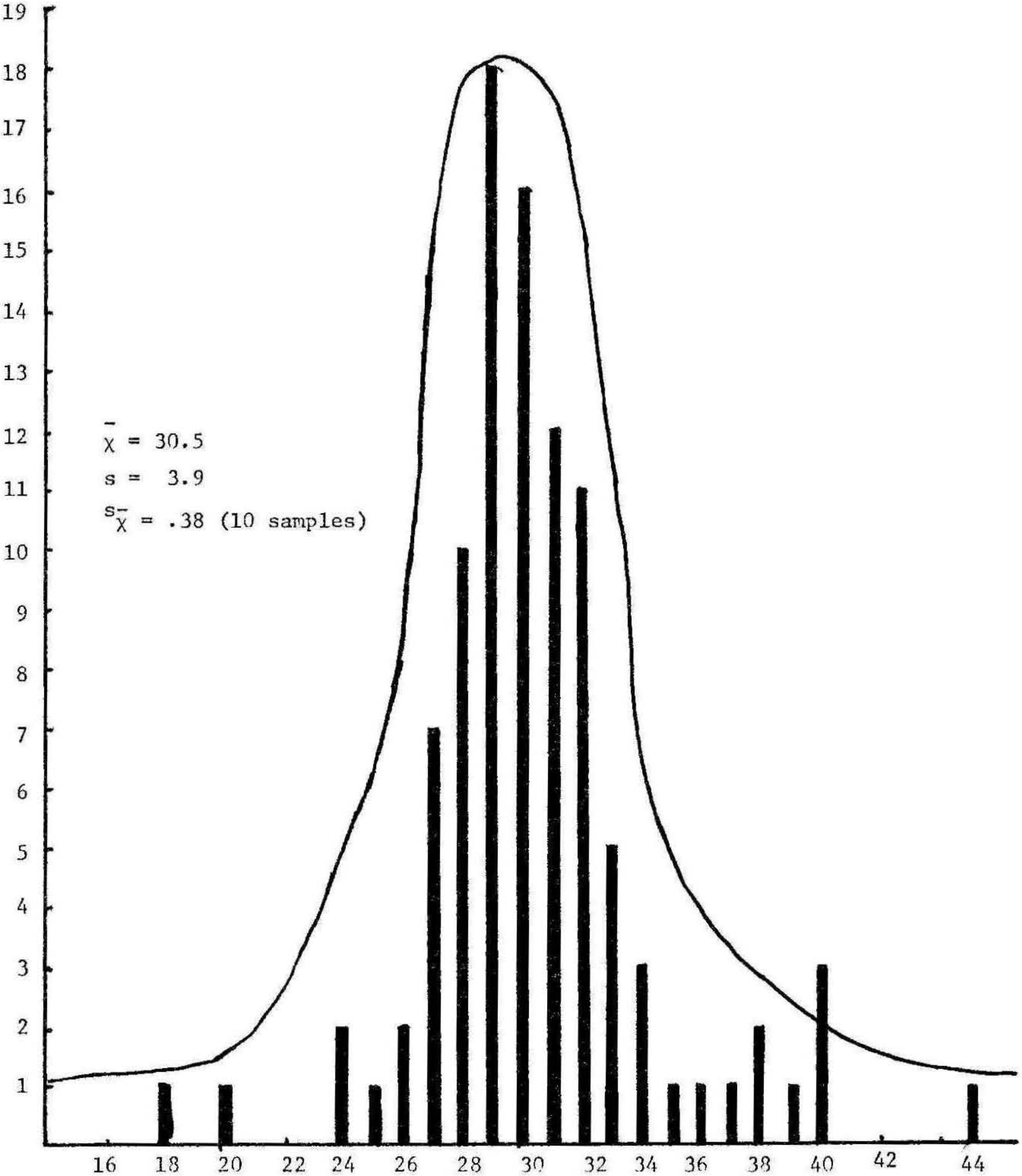

Figure 5.--Bar graph of frequencies of seedling densities in figure 4.

where Σ means to add together all the sample counts, x is a sample count, and n is the number of counts made. An example of a sample mean for 5 samples counts would be:

$$x = \frac{31+28+27+30+29}{5} = \frac{149}{5} = 29$$

The second calculation is for determining the sample standard deviation. This is a measure of the spread of the data or how variable it is. The sample standard deviation is computed as follows:

$$s = \sqrt{\frac{\Sigma X^2 - \frac{(\Sigma X)^2}{n}}{n-1}}$$

This formula is most simply described by using the data from the example of the mean above.

$$s = \sqrt{\frac{(31)^2 + (28)^2 + (27)^2 + (30)^2 + (29)^2 - \frac{(145)^2}{5}}{5 - 1}}$$

$$s = \sqrt{\frac{961 + 784 + 729 + 900 + 841 - 4205}{4}}$$

$$s = \sqrt{\frac{4215 - 4205}{4}} = \sqrt{\frac{10}{4}} = \sqrt{2.5} = 1.58$$

Figure 6 shows how the s, sample standard deviation, describes how variable the counts are. Within one standard deviation above and below the mean ($\pm$ s) 68 percent of all other observations will fall; 95 percent are within $\pm$ 2s; and 99 percent are within $\pm$ 3s of the mean.

The hypothetical nursery beds diagramed in figures 4 and 7 give us some idea of how this relates to nursery inventory. For easier discussion we can call these bed 1 and bed 2 respectively. The seedling counts for bed 1 are graphed in figure 5 and the counts for bed 2 are graphed in figure 8. The mean value, $\bar{x}$, for these two beds are close, however, the standard deviation is twice as large in bed 2 as it is in bed 1. The importance of this difference in standard deviation is this. In bed 1 our random samples might be all from one side of the distribution, but because it is more compact, the estimate of the mean would not be too greatly in error. With bed 2 and its larger standard

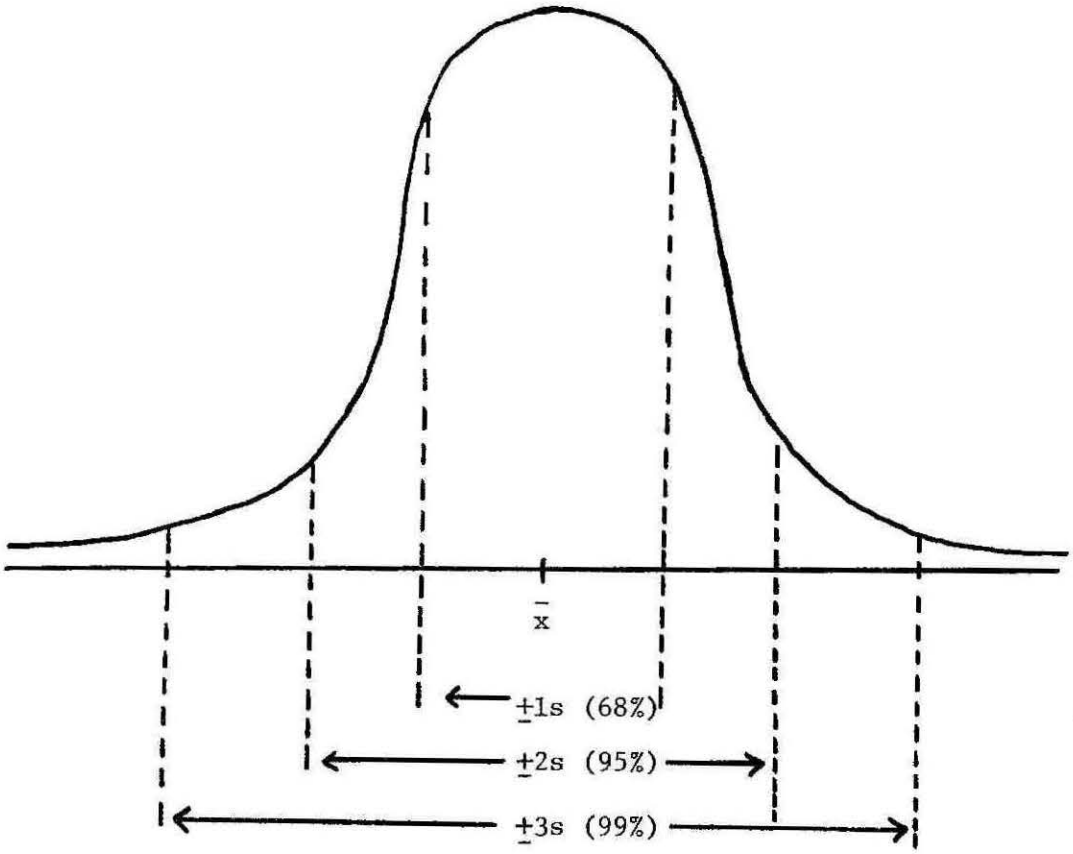

Figure 6.-- Structure of a standard normal distribution.

deviation we could be in greater error if our samples tended to come from mostly one side of the distribution. To compensate for this greater chance of error, we must take more samples. This at least reduces our chance for error.

The computation of the standard error of the mean, $s_{\bar{x}}$ is a way to describe with a number the effect we discussed in the last paragraph.

$$s = s/\sqrt{n}$$

where s is the sample standard error and n is the number of observations. For the example we used previously in this section we have that

$$s_{\bar{x}} = \frac{1.58}{\sqrt{n}} = \frac{1.58}{2.24} = .70$$

This $s_{\bar{x}}$ is some measure of how close a second measurement on the bed average would be if the boss were to check our work. For our simple example of five samples, we would expect that someone checking our work would have an $\bar{x}$ within 1.4 seedlings of our $\bar{x}$ of 29, 19 times of 20 checks.

10	38	45	28	30
20	25	40	29	29
30	27	28	30	35
28	32	30	32	15
29	15	31	31	19
32	18	32	27	23
35	20	28	29	29
60	29	28	30	40
55	32	29	32	42
40	34	30	32	35
35	40	40	31	32
31	50	32	30	33
29	60	18	27	18
32	53	19	28	22
34	40	15	31	27
40	32	31	29	32
48	29	29	33	37
31	31	32	40	31
32	40	30	41	29
28	30	29	35	28

Figure 7.--A diagram of a hypothetical nursery bed, with relatively high variability, showing the number of seedlings per square foot. Called bed 2 in the text.

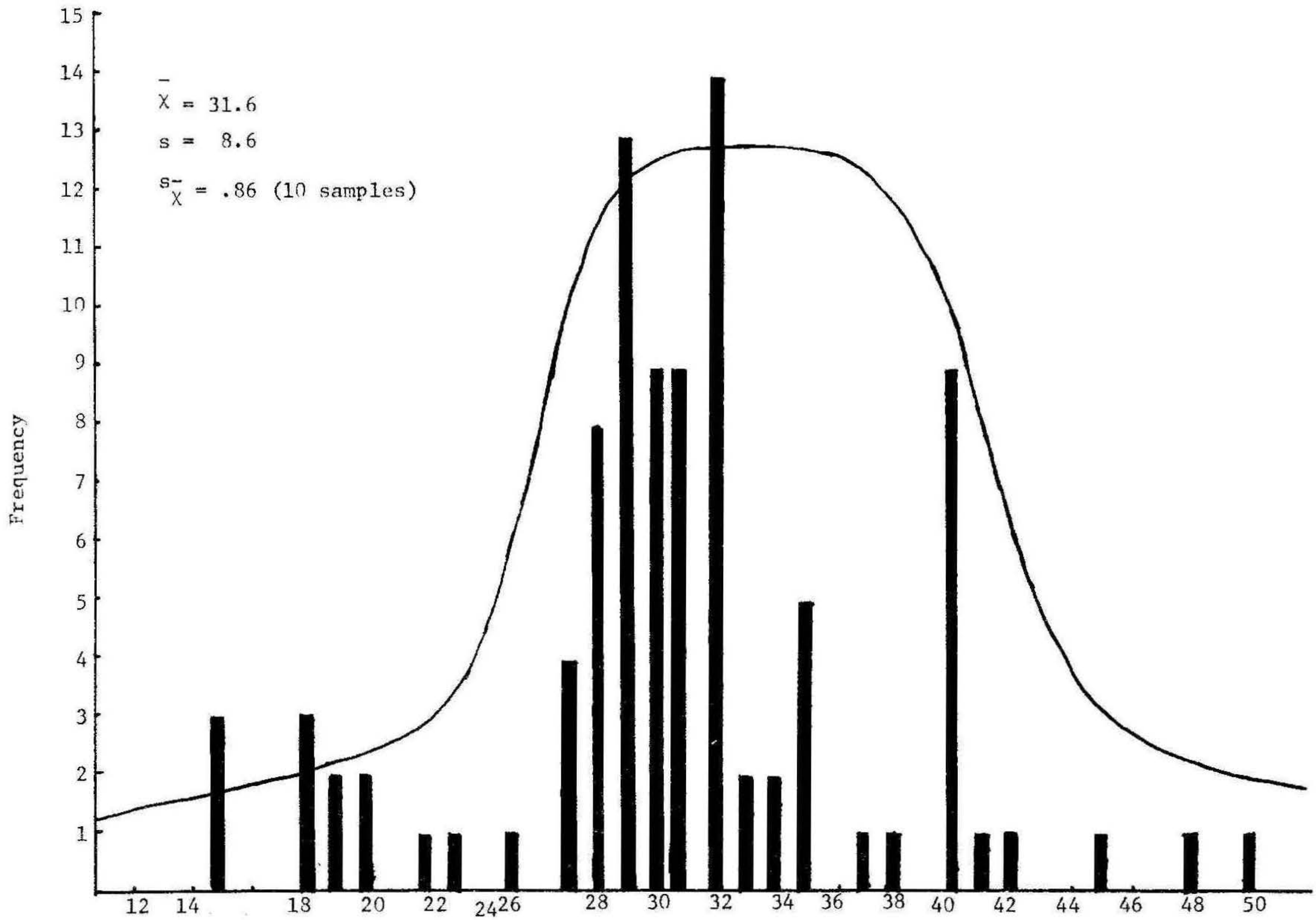

Figure 8.--Bar graph of frequencies of seedling densities in figure 4.

Our confidence that $\bar{x}$ is a precise estimate of the true average can be measured as follows:

$$\bar{x} \pm t_{.05,n} s_{\bar{x}} \quad .$$

Here $\bar{x}$,n and $s_{\bar{x}}$ are what we have already defined them as. The t is from a table of t values that can be found in most introductory statistic books. The .05 on the t is the error level. A .05 error level means that we will expect to be wrong only 1 time in 20 in the statements we make about what the average number of seedlings might actually be.

Continuing our simple example, our confidence interval will be as follows:

$$\bar{x} \pm t_{.05,4}\ s_{\bar{x}}$$

$$29 \pm 2.776\ (.70)$$

$$29 \pm 1.94$$

$$(27.1\ ,\ 30.9)$$

These computations lead us to say that we are wrong only 1 time in 20 when we say that the true average number of seedlings per square foot is between 27 and 31.

A simplified procedure for selecting a t value can be adopted if at least 10 samples are counted. This is because the change in the t value is relatively small when going from 10 samples to over 100 samples, especially when we consider how large the changes in standard error can be. Therefore, we can say that t will be 2.3 for an error level of 5 percent and will be 3.2 for an error level of 1 percent. Errors that result from using a constant t value are on the side of safety.

TYPES OF SAMPLING EXPLAINED

Systematic

Systematic sampling is the taking of a sample at fixed intervals, say every 20 feet, over the entire nursery bed. An example will be used to illustrate the procedure. We will adopt the sampling interval of every 20 feet. To start we randomly choose a number from 1 to 20. This can be done by drawing a number from a hat. Supposing the number is 7. Then we will measure in 7 feet from the end of the bed and make our count of seedlings on our 1 x 4 foot sample. The next sample will be taken at 27 from the end, the next at 47 and so forth. Choosing a number from the hat to tell us where to start is called making a random start. This is necessary if we want to use the statistics we discussed in the last chapter. The statistical calculations are very important because they are the only way to evaluate the precision of estimate, short of counting all seedlings.

Systematic sampling is somewhat easier to apply and relocate plots to verify previous counts. This is because of the regular intervals. However, we do not have the ability to improve the precision of our estimates as we have with random sampling.

Random Sampling

Plots are located by chance when sampling at random. Place 100 slips of paper, numbered 1 to 100, in a hat and thoroughly mix. Draw out as many slips of paper, one at a time, as there are samples to take. Supposing we desire to take 10 samples (this is a 10 percent sample) and the 10 slips of paper we draw have the numbers 8, 50, 3, 75, 42, 36, 19, 72, 56, 37. Then we measure in from the end of the bed 3, 8, 19, 36, 37, 42, 50, 72 and 75 feet and take a sample at each measured mark. Note that some plots are close together. This is to be expected because every lineal foot has an equal chance of being chosen. There is no problem in this unless the portion with more samples is noticeably different from the portion of the bed with few samples. In such cases we need to divide the bed into separate sampling units. More will be said about this in the last section. This system is a little more complicated to apply but offers the advantages of refining the estimate and minimizing the number of plots measured.

We will work through the application of random sampling on two hypothetical nursery beds. These two beds are diagramed in figures 4 and 7. Figure 9 is a data sheet that could be used to collect data and that we will use in our example. Figures 10 and 11 are worksheets that will be useful for computing and recording our estimates. In both figures 4 and 7 the circles indicate the first 10 samples taken, the squares the additional plots included to make the second estimate and the triangles the additional plots included to make the third estimate. The estimate number 1 for both beds in figures 10 and 11 was computed using the ten circled plots. Estimate 2 was then computed using the same 10 circled plots and the four plots marked with a square.

The mean and standard deviations in figure 10 were computed using a calculator with special functions to give the final answers directly without using the formulas of the previous section on variability. There are many relatively low cost machines that have these functions.

For estimate 1, bed 1, the mean number of seedlings per square foot is 31 or 12,400 seedlings in the whole bed. The standard error was 4.4 seedlings per square foot. The standard error of the mean is obtained by dividing the 4.4 by the square root of the number of samples which is 3.16. Therefore, the standard error of the mean is 1.39 seedlings per square foot. The 95 percent confidence interval for the mean is 31 ± 2.3 (1.39) or 31 ± 3.2. This confidence tells us that we are 95 percent certain in expecting the true average to be between 28 and 34 seedlings per square foot, or that the whole inventory in bed 1 is between 11,120 and 13,680 seedlings. If we are satisfied with being 95 percent sure we have between 11,120 and 13,680 seedlings we stop and go on to the next bed. By adding 4 more samples in making estimate 2 for bed 1, we narrowed the range in which the true average is expected to occur.

For bed 2, three estimates were made. With each estimate the average changed little. The standard error of the mean, however, dropped sharply by making the second estimate. The effect of this was to narrow the interval, by about one third, in which we expect to find the true average. In specific terms, our estimate of the number of trees in bed 2 can be expected to not be in error by more than 2,700 trees. And there is a 5 percent chance that this statement is incorrect. With estimate 2 we expect to be in error by no more than about 1,900

Bed 1			Bed 2		
Sample Number	Feet from end of bed	Seedlings per square foot	Sample Number	Feet from end of bed	Seedlings per square foot
1	3	38	1	3	30
2	10	32	2	15	34
3	33	28	3	29	32
4	43	32	4	33	60
5	51	30	5	36	32
6	73	27	6	50	30
7	83	40	7	64	32
8	93	29	8	67	29
9	95	28	9	76	29
10	98	30	10	90	35
11	24	34	11	16	40
12	27	32	12	44	30
13	68	30	13	46	32
14	79	27	14	97	37
15			15	9	55
16			16	54	19
17			17	69	32
18			18	78	40
19			19		
20			20		

Figure 9.--One possible worksheet for nursery inventory using random sampling.

Bed number	1		
Bed size	400		
Estimate number	1	2	
Number of samples (n)	10	14	
Mean($\bar{x}$) per sq. ft.	31	31	
Total	12,400	12,400	
Standard deviation (s)	4.4	3.9	
$s_{\bar{x}}$	1.39	1.05	
$2.3(s_{\bar{x}})$	3.2	2.4	
Confidence interval			
per sq. ft. low	27.8	28.6	
high	34.2	33.4	
Total bed low	11,120	11,440	
high	13,680	13,360	

Figure 10.--Worksheet for recording and computing estimates.

trees. The amount we can be off in our estimate has reduced because by using more samples the standard error of the mean was reduced. Estimate 3 failed to reduce the size of our confidence interval because the standard error increased slightly just by chance.

How many plots to count is an important question in random sampling. The cost of inventory is least when the fewest plots are counted, but this cost saving must be measured against the accuracy of the estimate. How accurate the estimate must be is the decision for the nurseryman.

Going back to figure 11, we see that our confidence interval for the mean for bed 2, estimate one, is from 10,880 seedlings to 16,320 seedlings.

Our estimate of the average, $\bar{x}$, is 13,600 seedlings. Therefore, if we promise this number of seedlings, we can be 95 percent sure that we would not be more than 2,720 seedlings short. In some cases there would be extra. In this case, we would not expect more than 2,720 extra seedlings. If we can live with

Bed number	2		
Bed size	400		
Estimate number	1	2	3
Number of samples (n)	10	14	18
Mean($\bar{x}$) per sq. ft.	34	34	35
total	13,600	13,600	14,000
Standard deviation (s)	9.3	8.0	9.5
$s_{\bar{x}}$	2.94	2.1	2.24
$2.3(s_{\bar{x}})$	6.8	4.8	5.2
Confidence interval			
per sq. ft. low	27.2	29.2	29.8
high	40.8	38.8	40.2
Total bed low	10,880	11,680	11,920
high	16,320	15,520	16,080

Figure 11.--Worksheet for recording and computing estimates.

the chance of being 2,720 seedlings short we can quit. If we have to be more certain, then we should take more samples as we did for estimate 2, bed 2. With this estimate, we are 95 percent certain that we will not be more than 1,920 short or over.

For bed two, we see that the confidence intervals on the average are smaller because of the lower variation in the bed. Therefore, the average estimate is used with greater confidence of being closer to the actual number of trees, the true average. A more uniform seedbed should be the aim of the nurseryman.

HISTORY PLOTS AND NURSERY INVENTORY

The primary purpose of history plots is for monitoring seedling growth and mortality. They are permanent sample plots located at random. What advantage do history plots offer over a general inspection of the seed beds? With a general inspection we can only make a guess of the amount of mortality and probably will not detect losses until an advanced stage. On a history plot, we know exactly how many trees are present, and can easily verify how many seedlings have died or are showing disease symptoms. In short, we can be more objective and specific in our determinations of crop survival or mortality. The early detection of mortality or above average survival can not be overemphasized if we consider how beneficial it will be to know that our survival is 10, 20 or even 30 percent below what we predicted. The advantage of history plots over spring inventory is that history plots represent less than 1 percent of the area, so they can be monitored rapidly.

History plots can also be used for inventory work. However, in this case extra random plots are taken in the general area of the history plot. The sample mean of these plots is calculated as well as the precision of the estimate of this mean. Enough extra samples need to be taken to give the desired precision just as we did in random sampling. Then for inventory purposes, we adjust the seedling count on the history plot up or down according to how it deviates from the mean of the extra plots. To this point, history plots are as much work as a random sampling inventory. The benefit will come in summer and fall inventory when only the history plots need to be measured.

A short example illustrates the procedure. The history plot has 30 seedlings per square foot and the extra plots 25 seedlings per square foot. For inventory purposes, then we will always reduce the count on the history plot by 1/6 or 17 percent. Adjustments are always made on a percentage basis.

To evaluate the percent cull factor, one half of the seedlings on the history plot is dug with a shovel and graded. The inventory count is reduced by the percent of culls. If there are 30 million seedlings and 10 percent culls, the plantable inventory would be 27 million.

CONTROLLING BED VARIATION AND MANAGEMENT OF THE NURSERY

Greater variation makes for greater problems in making accurate inventories and for keeping costs down. As we saw earlier, fewer samples were needed to obtain a desired confidence interval on the mean when the sample standard deviation was smaller. Fewer samples make for less work and, therefore, less cost. Therefore, controlling the variation is critical. One way to do this is to divide the nursery into parts that internally are uniform. Some examples of areas that would be internally uniform are areas of different soil types, areas prone to flooding, beds sown to one seed lot, and beds damaged by storms. These are types of variation which could be difficult or impossible to control. However, by recognizing where this variation exists, we can set boundary lines around the different areas and estimate a separate mean for each area.

There are practices that can reduce variation. These practices would include but not be limited to, working for uniform soil conditions, even water drainage, even application of pesticides, accurate seed sizing and sowing, use of high vigor seed and top pruning of seedlings. Because improved management gives lower variation, we can use our measures of variation as an objective way to evaluate our management practices. If variation is high we can expect that reducing variation will result in increased production and higher quality seedlings.

CONTROL OVER LIFTING AND PACKING

The purpose of a nursery inventory is to estimate the number of trees available for packing. Therefore, good control over the number of seedlings packed per bag or bundle is essential. A poor level of control at lifting can make even the most accurate of inventories meaningless. Whether seedlings are packed according to actual counts or by weight is not important. What is important is that someone has continuous responsibility to verify the counts and that a system exists to make corrections for errors.

Conducting an accurate inventory, controlling the variation, and maintaining control over packing require time, money and effort. Often it seems difficult to have enough of each to do all the jobs we are expected to do. Putting enough into inventory control is important to guide other practices and to maintain a good image for the nursery. In other words, a good inventory system can be indispensable in gaining maximum return from scarce resources and maintaining good relations with our customers and superiors who will supply resources to the nursery.