Exercise 2

Sampling Precision and Reliability

Howard Ferris
Division of Nematology
University of California, Davis 95616

1. <u>Objectives</u>
To demonstrate the relationship between precision and reliability of population assessment and the intensity of the sampling effort.

This exercise consists of two parts. All students should complete the first part, and the second should be carried out by those who are interested in a more in-depth analysis of the relationship between nematode populations (their density and distribution), the number of samples taken, and expected precision of the population estimates. The second part of the exercise will interest students who are motivated to do the suggested background reading, have some background in statistics, and have a special interest in problems related to sample size and precision of population estimates.

2. <u>Length of time required</u>
2 x 3hr laboratory periods, and some homework.
   Sampling: 2 hr
   Extraction: 2 hr
   Counting: 2 hr
   Analysis: 2 hr

3. <u>Procedure</u>

Part I:
a) Locate a field of 5-10 acres in size, either in a preplant fallow condition or with an established crop.
b) The class divides into groups, with two students per group.
c) Stratify (divide) the field, so that each group works on a portion of equal size. Note areas of homogeneous soil texture, soil moisture, previous cropping history, crop performance pattern, etc.
d) Each group samples a different (stratum) section of the field, removing 10 bulked samples to be analyzed separately for population density. Each sample should consist of at least 12 subsamples (cores), taken with a 1" diameter Oakfield tube to a depth of 12" below the soil surface. The 12 cores, or subsamples, of each sample unit are distributed across the whole stratum, similar to the zig-zag pattern diagrammed in Fig. 1.

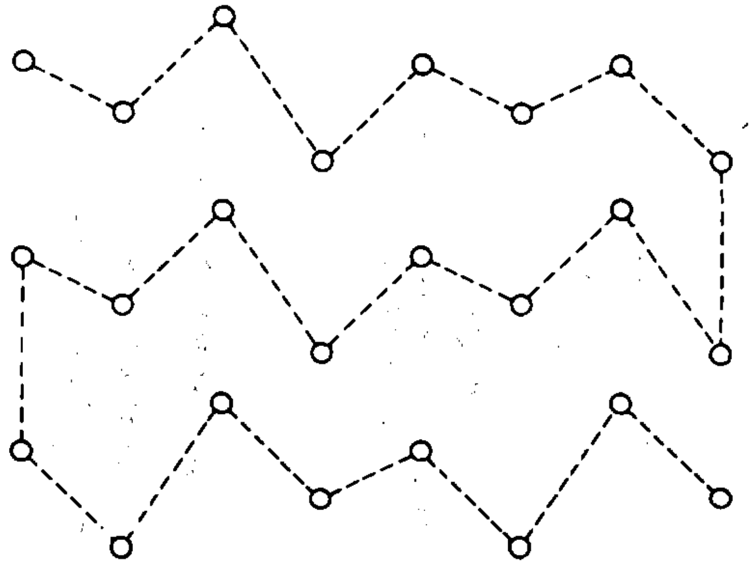

Fig. 1. Recommended pattern for collecting soil cores in a fallow field or established row or field crop.

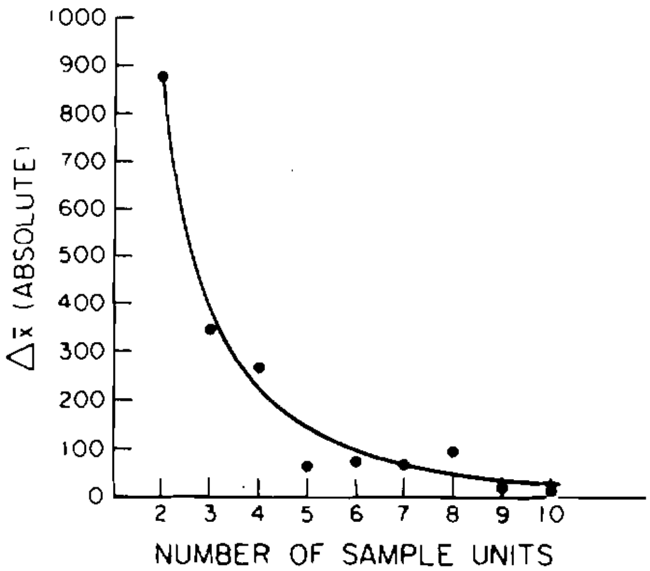

Fig. 2. Relationship between population assessment precision and number of sample units (data from example provided).

e) Alternatively, each group may work in the same stratum, but remove samples of different unit size (number of cores).
f) Return the soil samples to the laboratory and extract nematodes (this may be performed as the subject of laboratory exercise 1).
g) Count and identify the number of individuals of one or more genera of nematodes in each sample unit.
*Note: This presupposes some previous experience in nematode identification.
h) For your 10 sample units, determine the amount of

change in the population density estimate ($\Delta\bar{x}$) as the number of sample units increases. Graph the absolute value of the change in the population density estimate against number of sample units as in Fig. 2. Example of calculations:

Sample #	Population density Species A	Progressive mean ($\bar{x}$)	$\Delta\bar{X}$ Absolute
1	2108	2108	--
2	357	1233	875
3	2256	1574	341
4	510	1308	266
5	1622	1371	63
6	1805	1443	72
7	986	1378	65
8	642	1286	92
9	1431	1302	16
10	1215	1293	9

$\bar{x} = 1293$
$S^2 = 443816$

i) From the graph, determine the number of sample units at which increase in sampling intensity is not warranted by increase in assessment precision. Observe the lack of reliability of low numbers of sample units as an estimator. If different groups removed sample units of differing core number, observe the effect of sample unit on precision. Calculate an approximate cost per sample unit based on time involved in obtaining the data. Comment on the relationship between precision and cost.

Part II

a) For any sample unit size, calculate the mean and the variance of the population estimates for a single species. Use all 10 sample units in your calculations.

Background:

Spatial distributions of many species of plant-parasitic nematodes are often adequately described by the negative binomial model (Ferris, 1984). Populations occur in aggregates, and characteristic of this type of distribution, the variance of the population estimate is greater than the mean.

Based on this model, the number of samples required to estimate a population density with a specified level of precision varies directly with the density of the population, and can be determined according to the following equation:

$$n = t\alpha^2 \; (1/x + 1/k) d^2 \quad \ldots\ldots\ldots\ldots(i)$$

where n = the number of sample units required

$\bar{x}$ = the mean population density in the field

d = the acceptable deviation of the population estimate from its true value; if it is acceptable to measure the population within 25% of its true value, d = 0.25.

α = the proportion of predictions which will not fall within the desired confidence limits.

t = student's t-value at the required confidence level, with specified degrees of freedom (d.f.)

k = the dispersion parameter of the negative binomial. (This must be known before the equation can be solved).

To use this model to determine n, use your mean and variance estimates to estimate k, which is specific for the different species. It can be approximated by:

$$k = \bar{x}^2 \;/\; (S^2 - \bar{x}) \quad \ldots\ldots\ldots\ldots(ii)$$

From the field data set:

$$k = (1293)^2 \;/\; (443816 - 1293) = 3.78$$

Once k has been determined, you must select a value for "t" which enables you to calculate the sample size necessary to estimate $\bar{x}$ to within 0.25 of μ (the true mean population density). (Equation [i]).

Again, based on the field data set, with t for d.f.= ∞ (an infinitely large sample set and α = .05):

$$n = (1.96)^2 \; (1/1293 + 1/3.78) \;/\; (0.25)^2$$

$$n = 16$$

An initial value of an infinite number of d.f. is obviously unrealistic but it gives us a first approximation of n. We can improve this approximation by repeating the calculation using the earlier result as the d.f. for selecting a t-value.

Example: Where d.f. = n - 2 = 14, t.05 = 2.145.

$$n = (2.145)^2 (1/1293 + 1/3.78)/(0.25)^2$$

$$n = 20$$

This iterative process should be repeated until repeated estimates of n are very close.

This calculated n-value indicates the number of samples required to estimate a mean population density to within 25% of its true value at a 95% confidence level.

b) From background reading (McSorley and Parrado, 1982; Taylor, 1984) comment on the implications of the magnitude of k in terms of population distribution.

c) Explore and discuss the impact of required precision (d value) on the number of sample units for the species for which you obtained mean and variance data. Explore the effect of k value magnitude on the number of sample units required for a given level of precision.

d) Rewrite equation (i) as:

$$d = t\alpha\sqrt{(1/\bar{x} + 1/k)/n}$$

to determine the precision of population estimation associated with removing 1 or 2 sample units from the field.

USEFUL REFERENCES

Elliott, J. M. 1977. Some methods for the statistical analysis of samples of benthic invertebrates. Freshwater Biol. Assoc. Sci. Publ. 25. Ambleside, England.

Ferris, H. 1984. J. Nematol. 16:246-251.

Goodell, P. B. and H. Ferris. 1981. J. Nematol. 13:304-313.

McSorley, R. and J. L. Parrado. 1982. J. Nematol.14:522-529.

Taylor, L. R. 1984. Ann. Rev. Entomol. 29:321-357.