

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

Faculty Publications: Department of
Entomology

Entomology, Department of

October 1985

Use of Double Sample Plans in Insect Sampling with Reference to the Colorado Potato Beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae)

Jan P. Nyrop

New York State Agricultural Experiment Station, Geneva, N.Y.

Robert J. Wright

University of Nebraska-Lincoln, rwright2@unl.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/entomologyfacpub>



Part of the [Entomology Commons](#)

Nyrop, Jan P. and Wright, Robert J., "Use of Double Sample Plans in Insect Sampling with Reference to the Colorado Potato Beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae)" (1985). *Faculty Publications: Department of Entomology*. 106.

<https://digitalcommons.unl.edu/entomologyfacpub/106>

This Article is brought to you for free and open access by the Entomology, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Faculty Publications: Department of Entomology by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

Use of Double Sample Plans in Insect Sampling with Reference to the Colorado Potato Beetle, *Leptinotarsa decemlineata* (Coleoptera: Chrysomelidae)

JAN P. NYROP¹ AND ROBERT J. WRIGHT²

Environ. Entomol. 14: 644-649 (1985)

ABSTRACT When used in a pest control decision-making context, protocols for obtaining sample information on insect numbers can be evaluated via three properties: the operating characteristic (OC), the average sample number (ASN), and the availability of methods for calculating the OC and ASN. On the average, sequential sample plans have lower ASN's, and OC's are comparable to fixed sample-size procedures. The two most commonly used sequential procedures for insect sampling are Wald's probability ratio test and a test proposed by Iwao. Each of these methods has limitations. An alternative to sequential procedures is double sampling. Double sample plans developed for the Colorado potato beetle, *Leptinotarsa decemlineata* (Say), on Long Island, N. Y., have OC's comparable to a modified form of Iwao's sequential method and fixed sample size methods. They also have ASN's that at times are lower than those of the sequential method.

A BASIC TENET of pest management is that pesticides should be used to control pests only when needed. Decisions with respect to the need for control are often based on sample data from which an estimate of a population parameter is calculated and an inference about a population parameter is made. The set of all possible inferences map into a set of pest-control decisions. In this way, a decision is reached for any inference drawn.

As with all decision making under uncertainty, incorrect pest-control decisions can be made. One source of error arises from incorrect estimation and the consequent drawing of an incorrect inference. With no restriction on sample size, this source of error can be eliminated if estimators are unbiased. However, a tradeoff must usually be made between the cost of making an incorrect decision based on the sample data and the cost of data collection.

When used in a decision-making context, all sample plans can be evaluated via three properties. The first property is the operating characteristic (OC). Formally, this is defined in the context of hypothesis testing as the probability of accepting a null hypothesis given any simple hypothesis ($P[\text{accept } H_0 | H_i \text{ with } H_i; \mu = q]$ where μ is the population mean and q is a specific value of the mean). However, when used in decision making this is also a property of sample plans that yield parameter estimates (e.g., fixed sample size procedures) since these estimates are used to make decisions and one decision can be considered the null hypothesis. Ideally, sample plans minimize the probability of making incorrect decisions.

The second property is the average number of samples (ASN) required to terminate the sampling process given any true population parameter. Obviously, it is desirable to keep the ASN as low as possible subject to constraints of the OC.

The final property is the availability of methods for easily determining the OC and ASN. Without such methods it is impossible to judge the relative worth of sample plans in a decision-making context.

In insect sampling, most sample plans based on hypothesis testing are sequential. Wald's sequential probability ratio test (SPRT) (Wald 1947) and a test proposed by Iwao (1975) are used most often. In general, sequential hypothesis tests have lower ASN's, and OC's are comparable to fixed sample size procedures based on estimation. However, the SPRT and Iwao's procedure have limitations.

Use of the SPRT assumes that sample data can be described by a probability model with one unknown parameter. This assumption cannot often be met and its violation can lead to unpredictable changes in the OC and ASN.

Iwao's procedure is founded on a mean variance relationship

$$\sigma^2 = g(\mu) \quad (1)$$

where g represents some function. Such a relationship can often be found. However, the OC is difficult to specify a priori and both the OC and ASN must be determined via simulation once a sample protocol has been constructed (Nyrop and Simons 1984).

Double sampling is an alternative to sequential hypothesis tests with somewhat different properties that is not encumbered by some of the previously mentioned difficulties. It has been used extensively in quality control (Cowden 1957). In this paper we review the basis for constructing double

¹ Dept. of Entomology, New York State Agric. Exp. Stn., Geneva, NY 14456.

² Dept. of Entomology, Cornell Univ., L. I. Hort. Res. Lab, 39 Sound Avenue, Riverhead, NY 11901.

Table 1. Dispersion of CPB in commercial potato fields as described by Taylor's (1961) power law ($\sigma^2 = a\mu^b$), Long Island, N.Y., 1982-1983

CPB life stage	Year	$\log_{10}a \pm 95\% \text{ CL}$	$b \pm 95\% \text{ CL}$	r^2	n
Adults	1982	0.18 ± 0.002	1.11 ± 0.002	0.953	213
	1983	0.18 ± 0.002	1.11 ± 0.003	0.956	165
	Combined	0.18 ± 0.002	1.11 ± 0.001	0.957	378
Egg masses	1982	0.18 ± 0.003	1.11 ± 0.003	0.934	217
	1983	0.15 ± 0.002	1.08 ± 0.002	0.968	167
	Combined	0.17 ± 0.001	1.09 ± 0.001	0.954	384
Small larvae (first-second instar)	1982	1.03 ± 0.003	1.66 ± 0.004	0.954	173
	1983	1.00 ± 0.004	1.54 ± 0.006	0.939	140
	Combined	1.01 ± 0.002	1.60 ± 0.002	0.947	313
Large larvae (third-fourth instar)	1982	0.58 ± 0.003	1.43 ± 0.004	0.940	177
	1983	0.55 ± 0.004	1.35 ± 0.006	0.903	144
	Combined	0.57 ± 0.002	1.40 ± 0.002	0.927	321

sample plans and illustrate their use for sampling Colorado potato beetles (CPB), *Leptinotarsa decemlineata* (Say). We compare this double sample plan via the OC and ASN to plans based on a modification of Iwao's sequential hypothesis testing procedure and a fixed sample size estimation plan.

Materials and Methods

Basis for Double Sample Plans. Two hypotheses concerning the population parameter of interest are constructed. In most cases, this parameter is a mean density (μ) and the hypotheses are;

$$\begin{aligned} H_0: \mu \leq D \\ H_1: \mu > D \end{aligned} \tag{2}$$

A sample of n_1 observations is taken. If the mean (m_1) of this sample is less than or equal to some A , H_0 is accepted. A satisfies the condition $A \leq D$. If $m_1 \geq R$, H_0 is rejected. R satisfies the condition $R \geq D$. The mean (m_1) is compared with A and R as opposed to D to guard against incorrect classification of the population when an estimate of μ has been obtained from n_1 observations. If $A < m_1 < R$, another sample of n_2 observations is taken and a mean (m_2) is computed. If $([n_1m_1 + n_2m_2]/[n_1 + n_2]) \leq D$, H_0 is accepted. Otherwise H_0 is rejected.

The OC is defined as $P\{\text{accept } H_0 | H_1 \text{ true}\}$. For a double sample plan sampling a continuous variate it is

$$\begin{aligned} B_i = P\{m_1 \leq A\} \\ + \int_A^R f(m_1)P\{m_2 \leq (D(n_1 + n_2) \\ - n_1m_1/n_2)\} dm_1 \end{aligned} \tag{3}$$

where $f(\cdot)$ is a probability density function. The computational form of equation 3 is

$$\begin{aligned} B_i = \int_{-\infty}^A f(m_1) dm_1 \\ + \int_A^R f(m_1)F\{D(n_1 + n_2) \\ - (n_1m_1)/n_2\} dm_1 \end{aligned} \tag{4}$$

where $F(\cdot)$ is a cumulative probability function. The ASN for a double sample plan with H_1 true is

$$\begin{aligned} ASN_1 = n_1 + n_2P(A \leq m_1 \leq R) \\ = n_1 + n_2(F[R] - F[A]) \end{aligned} \tag{5}$$

Two difficulties arise when developing a double sample protocol. First, choosing A , R , n_1 , and n_2 to produce a particular OC and ASN is not straightforward. Second, $f(\cdot)$ is often unknown.

The second difficulty may be overcome by making n_1 large enough to assume that the Central Limit Theorem applies to m_1 . Then f is a normal distribution with mean μ and variance σ^2/n_1 . The variance is also usually unknown. However, it can

Table 2. Dispersion of CPB in commercial potato fields as described by Iwao's mean crowding (m) regression ($m = a + b\mu$), Long Island, N.Y., 1982-1983

CPB life stage	Year	$a \pm 95\% \text{ CL}$	$b \pm 95\% \text{ CL}$	r^2	n
Adults	1982	0.156 ± 0.008	1.51 ± 0.023	0.265	213
	1983	0.169 ± 0.006	1.47 ± 0.014	0.619	165
Egg masses	1982	0.161 ± 0.007	1.68 ± 0.025	0.274	217
	1983	0.174 ± 0.007	1.30 ± 0.012	0.634	167
Small larvae (first-second instar)	1982	5.47 ± 0.122	4.98 ± 0.057	0.494	173
	1983	7.73 ± 0.164	3.13 ± 0.062	0.334	138
Large larvae (third-fourth instar)	1982	0.626 ± 0.038	3.32 ± 0.038	0.488	177
	1983	1.16 ± 0.040	2.51 ± 0.042	0.399	144

Table 3. Correlations between sample means of CPB and the normal scores of these means

Life stage	Density	r^2 Between sample means and normal scores of means		
		$n = 15$	$n = 20$	$n = 25$
Adult	Low	0.986	0.979	0.978
	Medium	0.996	0.991	0.998
	High	0.990	0.993	0.996
Small larvae	Low	0.883	0.921	0.933
	Medium	0.948	0.967	0.974
	High	0.981	0.985	0.992
Large larvae	Low	0.847	0.882	0.931
	Medium	0.985	0.978	0.988
	High	0.985	0.989	0.994

For each life stage and density, 200 sample means were generated for each n from a population of counts ($n = 300$) stored on a computer. The densities low, medium, and high refer to insects per 50 vines according to the following scheme: Adults: low, <15; medium, 15–25; high, >25. Small larvae: low, <75; medium, 75–200; high, >200. Large larvae: low, <30; medium, 30–75; high, >75.

often be approximated in terms of the mean as in equation 1.

Double Sample Plans for CPB. Double sample plans for three stages of the CPB were developed for use in a pilot integrated pest management (IPM) program on Long Island, N.Y. This was done in two stages. First, data were collected with which to deduce a variance–mean relationship as in equation 1 and to formulate hypotheses about mean densities for use as action thresholds. Second, the OC and ASN for double sample plans were computed and compared to those for other sample plans.

Samples of CPB were taken at least weekly from June through August from 12 potato fields in 1982 and 16 fields in 1983. In both years, a single potato vine was the observation unit. In 1982, 80 vines per field were sampled on each visit. Vines were selected in a nested fashion; four vines were selected randomly at each of 20 randomly selected sites in a field. In 1983, 50 vines (5 vines per site, 10 sites per field) were sampled. All aboveground CPB life stages were counted and categorized into

the following classes: adults, egg masses, small (first and second instars), or large (third and fourth instars) larvae.

Average intracluster correlation coefficients (Jessen 1978) were computed from 20 sets of 1983 data for each of the life stages. These average coefficients were all slightly greater than zero (adults: 0.135; egg masses: 0.056; small larvae: 0.041; large larvae: 0.160). The correlation coefficient is a measure of the efficiency of a sample unit when efficiency is measured in terms of the sample variance. If the coefficient is ≤ 0 , cluster (groups of vines per site) sampling is more efficient. Otherwise, random sampling of the individual observation units is better. For CPB, random sampling of individual vines is slightly more efficient in terms of the variance. However, in terms of time spent sampling, we have found cluster sampling more efficient. The mean time needed to sample a single vine is 26 s (SD = 6.1 s, $n = 100$) and the average time needed to travel between sample sites is 37 s (SD = 13.0 s, $n = 97$). To minimize sampling time, all observations should be taken in one sample location. Cluster sampling is a compromise between minimizing the variance of the population estimate and the time required to sample the population.

Results

Estimates of means and variances for each CPB class were fitted by linear regression to equation 1 by Taylor's power law (Taylor 1961) and Iwao's mean-density/mean-crowding relationship (Iwao 1968). Parameters for these relationships were estimated for each of the 2 years and for both years combined (Tables 1 and 2). The parameters estimated for Taylor's relationship are similar to those reported by Harcourt (1963) and Logan (1981).

Taylor's power law provided a better description of the relationship between the mean and variance than did Iwao's method. With Taylor's power law, the parameter estimates for each year were generally consistent with each other. Hence, parameter estimates used to develop the sample

Table 4. Parameters used to develop double sample plans for three life stages of the CPB

Life stage	Action threshold (D)	Population mean (A & R)	Probability of error	Irreversible decision
Adults	0.5/vine	0.3	0.057	
		0.9	0.044	
Small larvae	4.0/vine	2.1	0.051	Always > threshold
		8.0		
Large larvae	1.5/vine	0.9	0.047	
		2.7	0.061	

Small larvae are first and second instars and large larvae are third and fourth instars. The action threshold is the density of beetles for which control should be initiated. The column titled population mean is the mean (μ) for which one of two conditions is met. Either this is the mean for which the probability of classifying the population incorrectly (probability of error) is ca. 0.05 when 25 samples are taken, or the sample estimate of m for which a different decision with respect to the action threshold cannot be arrived at given another 25 samples. Probabilities of error were computed using a normal probability model. The action threshold is the parameter D and the population means are the parameters A and R of a double sample plan.

Table 5. Incidence of CPB above action thresholds in commercial potato fields, Long Island, N.Y.

CPB Stage	1982		1983	
	%	<i>n</i>	%	<i>n</i>
Adults	12.7	213	24.8	165
Small larvae	6.4	173	13.6	140
Large larvae	8.5	177	12.5	144

The action thresholds per vine are adults, 0.5; small larvae, 4.0; and large larvae, 1.5.

plans were based on the pooled 1982 and 1983 data.

On Long Island, CPB control decisions are based on counts of adults, small larvae, and large larvae. Action thresholds for these life stages were developed based on correlations of damage and CPB densities. Although these thresholds are undoubtedly conservative, they are useful as a first approximation until better thresholds are available. The thresholds for the different life stages are as follows: adult, 0.5 per vine; small larvae, 4.0 per vine; and large larvae, 1.5 per vine. Based on these thresholds, the null and alternate hypotheses concerning the mean density for each lifestage were formulated by substituting the action threshold for *D* in equation 2. Acceptance of the null hypothesis leads to a no-control decision.

Double sample plans were developed for each life stage in the following manner. Correlations between means and the normal scores of these means were generated to test the assumption of normally distributed sample means based on sample sizes of 15, 20, and 25 observations (Table 3). This was done by generating 200 random means for each sample size from the raw CPB count data and then determining the normal scores of these means. Based on these correlations, *n*₁ was set to 25 for each life stage. Based on the OC for a fixed sample size procedure, the cost of making an incorrect decision, and the cost of collecting sample data, it was decided that a maximum of 50 samples would be taken. Hence, *n*₂ was set to 25. The probability of making an incorrect decision based on 25 samples ($P[m_1 > D | \mu \leq D]$ or $P[m_1 < D | \mu \geq D]$) was computed using a normal model for a set of population means (μ) for each life stage. The parameters *A* and *R* were then selected as either the densities (μ) for which these probabilities were ca. 0.05, or as the sample estimate of μ_1 for which it was impossible to reach a different decision given another *n*₂ samples. Note that in the first case we are determining the μ for which these conditions are met and then substituting these values for *A* and *R*. The *n*₁ sample estimates are then compared with these values. This information is summarized in Table 4 for each life stage.

The OC and ASN for the double sample plans were computed using equations 4 and 6. The functions *f*() and *F*() in these equations are normal density and normal cumulative probability func-

tions. In equation 4, each was approximated with a polynomial (equations 26.2.18 and 26.2.20 in Abramowitz and Stegun [1972]).

Discussion

The three double sample plans developed were compared to sequential plans developed with a modified version of Iwao's (1975) procedure and with a fixed sample size of 50 observations. Iwao's procedure was modified two ways. First, 25 samples were drawn from the population before comparing the sample observations to the sequential stop boundaries. This was done to ensure that the count of total CPB (*T*) found in 25 samples could be approximately described with a normal density function. Commencing with the 25th sample (*n* =

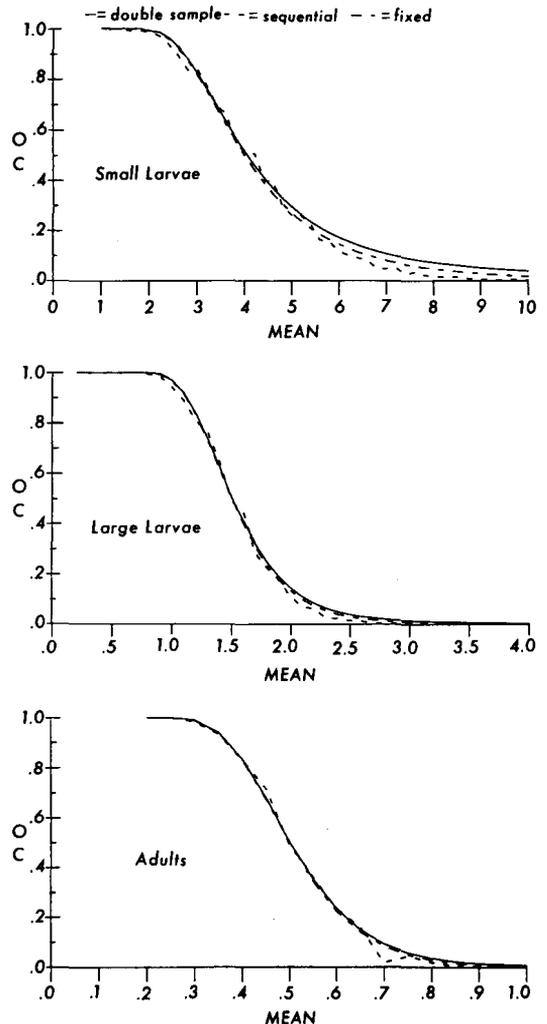


Fig. 1. OC of double sample plans, sequential sample plans, and fixed sample plans (*n* = 50) for three life stages of CPB. Small larvae are first and second instars and large larvae are third and fourth instars.

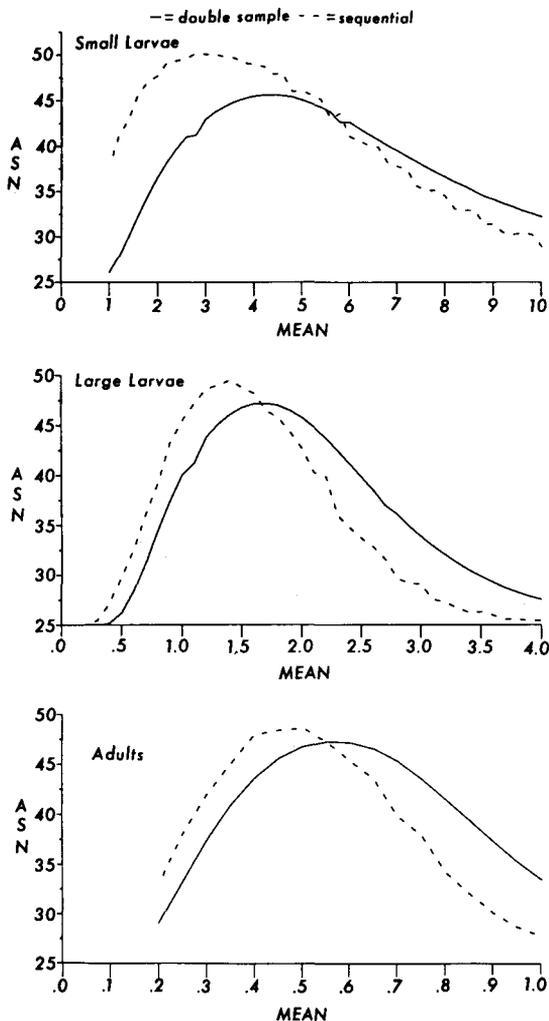


Fig. 2. ASN of double sample plans and sequential sample plans for three life stages of CPB. Small larvae are first and second instars and large larvae are third and fourth instars.

25), T was compared with two functions. If $T \leq D - 1.96(ng[D])^{1/2}$, H_0 was accepted. If $T \geq D + 1.96(ng[D])^{1/2}$, H_0 was rejected. Otherwise, another sample was taken until 50 samples had been taken. If this last criterion was met, a mean based on 50 samples was computed and compared with D . If the mean was $\leq D$, H_0 was accepted. Otherwise, it was rejected. The limit of 50 samples and decision making based on a sample mean constituted the second modification.

For the fixed sample size plan, a mean based on 50 samples was compared with D . If this mean was $\leq D$, H_0 was accepted. Otherwise, it was rejected.

The OC and ASN curves for the sequential procedure were generated by simulating 2,000 samples for a set of means from each life stage class.

Individual counts of CPB were generated from a negative binomial distribution with k specified as

$$k = \mu^2 / (\sigma^2 - \mu^2). \tag{7}$$

The OC for the fixed sample size procedure was calculated for any μ_i as

$$B_i = F(D). \tag{8}$$

The OC and ASN for each sample plan are shown in Fig. 1 and 2. The OC's for each plan are similar. When mean densities exceed the action threshold, the sequential and fixed sample size procedures are slightly better than the double sample plan. This is most noticeable for the small larvae. When the mean densities are less than the action threshold, the double sample plan performs slightly better than the sequential procedure. In terms of the ASN, the double sample plan requires fewer samples, on the average, than the sequential procedure with lower densities and the converse is true at higher densities. Therefore, if densities less than the action threshold are encountered most often, the double sample plan will be more parsimonious with respect to sample size. The reverse will be true if densities greater than the action threshold are encountered most frequently.

Data from 1982 and 1983 indicate that CPB densities are below the action threshold more often than they are above it (Table 5). Provided that this relationship does not change dramatically in future years, the double sample plan will be a more powerful and parsimonious sampling procedure to use. In 1984, the sample plan was used in a pilot IPM program on Long Island. In 65% of the fields scouted ($n = 368$) a decision was reached after 25 vines were sampled. This indicates that the proportions listed in Table 5 did not greatly change in 1984.

Double sample plans provide an alternative to sequential and fixed sample size plans when sampling for decision making. For CPB, double sample plans perform as well in terms of the OC as a fixed sample size procedure and a modified version of Iwao's (1975) sequential procedure. The double sample plans were always better in terms of the ASN than the fixed sample size procedure and, depending on the mean density of the CPB population, slightly better or slightly worse than the sequential procedure. Providing that the Central Limit Theorem holds for the means computed after the first sample of n_1 observations and that a good prediction of the variance can be made based on the mean, the OC and ASN of double sample plans are easily computed. Unfortunately, specifying the parameters of a double sample protocol to achieve a desired OC is not easily done. The method we employed to develop double sample plans for the CPB is general and provides a first step in designing double sample plans with a desired OC. Further changes in the OC and ASN must be made via trial and error changes in the

parameters of the plan. Unfortunately, no easy to follow formula is available for relating a particular change in the OC or ASN to changes in any one or combination of these parameters. The same is, of course, true with Iwao's (1975) sequential procedure. The OC and ASN for double sample plans are easier and less time-consuming (on a computer) to calculate than those for Iwao's procedure.

Acknowledgment

We thank L. Ulrevich, D. Kain, and D. Gilrein for data collection and assistance in data analysis. We also thank J. Barnard and A. Sawyer for reviews of an early version of this manuscript. The comments and suggestions of two anonymous reviewers were most helpful. This work was supported by the Cornell Univ., College of Agric. and Life Sciences, Office of Research and the New York State IPM Program.

References Cited

- Abramowitz, M., and I. A. Stegun.** 1972. Handbook of mathematical functions with formulas, graphs, and mathematical tables. U.S. Dept. Comm. Natl. Bur. Stand. Appl. Math. Ser. No. 55.
- Cowden, D. J.** 1957. Statistical methods in quality control. Prentice-Hall, Englewood Cliffs, N.J.
- Harcourt, D. G.** 1963. Population dynamics of *Lepidoptarsa decemlineata* (Say) in eastern Ontario. 1. Spatial pattern and transformation of field counts. Can. Entomol. 95: 813-820.
- Iwao, S.** 1968. A new regression method for analyzing the aggregation pattern of animal populations. Res. Popul. Ecol. 10: 1-20.
1975. A new method of sequential sampling to classify populations relative to a critical density. Res. Popul. Ecol. 16: 281-288.
- Jessen, R. J.** 1978. Statistical survey techniques. John Wiley, New York.
- Logan, P. A.** 1981. Estimating and projecting Colorado potato beetle density and potato yield loss, pp. 105-118. In J. Lashomb and R. Casagrande [eds.], Advances in potato pest management. Hutchison Ross, Stroudsburg, Pa.
- Nyrop, J. P., and G. A. Simmons.** 1984. Errors incurred when using Iwao's sequential decision rule in insect sampling. Environ. Entomol. 13: 1459-1465.
- Taylor, L. R.** 1961. Aggregation, variance, and the mean. Nature (London) 189: 732-735.
- Wald, A.** 1947. Sequential analysis. John Wiley, New York.

Received for publication 19 December 1984; accepted 2 July 1985.