This is to certify that the

thesis entitled

Comparison of Four Sequential Sampling Plans Applied to Forest Tent Caterpillar Eggs on Sugar Maple Branches

presented by

Jan P. Nyrop

has been accepted towards fulfillment of the requirements for

Master degree in Entomology

Gary C. Simmons
Major professor

Date August 10, 1979

COMPARISON OF FOUR SEQUENTIAL SAMPLING PLANS APPLIED TO FOREST TENT CATERPILLAR EGGS ON SUGAR MAPLE BRANCHES

By

Jan P. Nyrop

A THESIS

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

MASTER OF SCIENCE

Department of Entomology

1979

ABSTRACT

COMPARISON OF FOUR SEQUENTIAL SAMPLING
PLANS APPLIED TO FOREST TENT CATERPILLAR
EGGS ON SUGAR MAPLE BRANCHES

By

Jan P. Nyrop

Sequential sampling is a valuable tool for classifying population density. To date, all applications in insect sampling have used Wald's Sequential Probability Ratio Test (SPRT). Use of the procedure necessitates knowing the distribution of the underlying population and requires that the distribution be constant in time and space. When these assumptions are not met sequential t-tests and a new sequential test proposed by Iwao provide alternatives. Four sequential procedures were compared through simulation. These were the SPRT, Iwao's test and two sequential t-tests proposed by Barnard and Fowler and O'Regan. Forest tent caterpillar (Malacosoma disstria) egg band sampling was used as a test case. Though not universally true, in this instance SPRT was robust to changes in k of the negative binomial distribution. Fowler's and O'Regan's t-test and Iwao's procedure were comparable to the SPRT. Fowler's and O'Regan's t-test required the least information about the population distribution for construction.

ACKNOWLEDGMENTS

Many people have contributed both directly and indirectly to my thesis and I extend my thanks to them:

To my advisor, Dr. Gary Simmons, who first fired my interest in entomology and has since provided much support, professionalism and friendship.

To my committee, Drs. Tom Edens, Dean Haynes and Henry Webster who contributed greatly during my graduate work.

To Jeri, for her encouragement.

This work was supported in part by a grant from the Michigan Department of Natural Resources, Forest Management Division.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE PAGE

INTRODUCTION

Biological monitoring is an important component of any successful pest management program. Often, it is only necessary to classify an arthropod population as above or below a specified density. In such cases sequential hypothesis testing provides an attractive alternative to sampling schemes which employ a fixed number of observations. Fixed sample size methods are invariably inadequate at low population densities and excessive at high densities. With sequential procedures sample size is dependent on the outcome of each successive observation. These tests, on the average, require fewer observations than do equally reliable tests based on fixed sample size procedures. For this reason, they are attractive sampling schemes when cost and time efficiency are important.

To date, all applications of sequential hypothesis testing in insect sampling have employed Wald's Sequential Probability Ratio Test (SPRT) (Wald 1947). Pieters (1978) provides an extensive list of insect species for which such sampling schemes have been developed. Use of the SPRT requires that the underlying population distribution and variance be known. However, these parameters are often unknown and the population distribution upon which an SPRT is developed may change with changes in the density, quality and age of the population and spatial and temporal changes in the environment. Such changes will affect the results obtained using SPRT.

Alternatives to the SPRT exist. These are sequential t-tests and a new sequential procedure based on the regression of Lloyd's mean crowding on mean density (Iwao 1975). The latter requires an estimation of the population distribution based on the aforementioned regression, however, this relationship is reported to be stable in time and space

(Iwao and Kuno 1971). Sequential t-tests assume the population is approximately normally distributed.

The purpose of this paper is to compare, through simulation, four different sequential procedures. Forest tent caterpillar (Malacasoma disstria) (FTC) egg band sampling is used as a test case. The four sequential procedures investigated are the SPRT, Barnard's sequential t-test (BTTEST) (Barnard 1952), Fowler's and O'Regan's truncated sequential t-test (FTTEST) (Fowler and O'Regan 1974) and Iwao's sequential test (ITEST) (Iwao 1975). Simulation, the process of conducting experiments on a model, as opposed to attempting the experiment with the real system, provides an ideal tool for this evaluation. The result provides the reader with a method to weigh relative merits of each test and facilitate a choice between tests.

Development follows four parts: 1) A detailed description of the SPRT and a less intense outline of the other sequential procedures and problems in their use; 2) a description of the model used to investigate these problems and compare tests; 3) experiments conducted with the model and their results; 4) discussion and conclusion.

Wald's Sequential Probability Ratio Test

A. Maximum Likelihood Estimation

The SPRT is defined as the ratio of the probabiltiy of obtaining a given set of observations if an alternate hypothesis is true to the probability of obtaining the same set of observations if the null hypothesis is true. Initially, two hypotheses (H_0, H_1) of the actual population parameter (θ) are established. Consecutive samples are then examined and evaluated until cumulative results dictate acceptance of one of the hypotheses. The SPRT is based on the maximum likelihood

which is now introduced through illustration.

Assume x is distributed binomially with probability density function (p.d.f.);

$$f(x,p) = p^{x}(1-p)^{1-x} \quad \text{where } x = 0,1 \text{ and } 0 \leq p \leq 1.$$

We wish to find an estimation $u(X_1,X_2,X_3,\ldots,X_n)$ such that $u(x_1,x_2,x_3,\ldots,x_n)$ is a good estimate of p where $x_1, x_2, x_3,\ldots, x_n$ are observed. The probability that $X_1, X_2, X_3,\ldots, X_n$ takes on these particular values is;

$$P(X_1=x_1, X_2=x_2, X_3=x_3,\ldots, X_n=x_n)$$

which equals;

$$\prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

This is the joint p.d.f. of $x_1, x_2, x_3,\ldots, x_n$. An estimate of p may be arrived at by regarding the p.d.f. as a function of p and finding the value of p which maximizes it. In other words, we wish to find the p value most likely to have produced these values. This is the likelihood function $L(p; x_1,x_2,x_3,\ldots,x_n)$ and will henceforth be designated as $L(\theta)$. Instead of finding the value p which maximizes $L(\theta)$ it is easier to find p which maximizes ln $L(\theta)$. The maximum is found by taking the first derivative and setting it equal to zero. Calculation will show the solution is;

$$1/n\ (\Sigma x_i)$$

or the mean as expected.

Neyman and Pearson (1928) suggested that a useful criterion for testing hypotheses is provided by the likelihood ratio;

$$\lambda = \frac{L(\theta) \text{ when } \theta = \theta_0}{L(\theta) \text{ when } \theta = \theta_1}$$

The hypothesis H_0; $\theta = \theta_0$ will be accepted when λ is large and H_1; $\theta = \theta_1$

accepted when λ is small. Values of λ may be selected to assure a specified level of α; the probability of accepting H_1 when H_0 is true and β; the probability of accepting H_0 when H_1 is true (Type I and II errors respectively). Wald (1947) applied this criterion in developing the SPRT. Approximate values of λ are calculated using the following logic.

Suppose a large number of sequential tests are made. Those which terminate with acceptance H_0 have a likelihood ratio equal to or slightly greater than λ_0. Those terminating with acceptance of H_1 have a likelihood ratio slightly less than λ_1. Consider the group of samples with likelihood ratio greater than λ_0. The probability that the sample really originated from the population with parameter θ_0 is λ_0 times as great as the probability that it originated from a population with parameter θ_1. This is true for every sample, thus the total chance of obtaining a sample of this sort is λ_0 times as large when H_0 is true as when H_1 is true. We desire the chance of getting such a sample from a population where $\theta = \theta_0$ to be $1 - \alpha$ and the chance to be β for $\theta = \theta_1$. Consequently, in order to satisfy both conditions λ_0 must equal $(1-\alpha)/\beta$. Using similar reasoning λ_1 must equal $\alpha/(1-\beta)$.

In practice, the likelihood function is solved for the total units found in Q observations. At each sample this total (T) is compared to a function of λ_0 and λ_1 and decisions made:

$T \geq f(\lambda_0)$ accept H_0

$T \leq f(\lambda_1)$ reject H_0

$f(\lambda_0) \geq T \geq f(\lambda_1)$ continue sampling

An illustration is given assuming a normal distribution:

The probability of a single observation from a normal distribution is;

$$\frac{1}{\sigma(2\pi)^{\frac{1}{2}}} \exp\ [-\frac{(x_i-u_0)^2}{2\sigma^2}\ ]$$

The probability of a sample of n independent observations x_1, x_2, x_3, ..., x_n is the product of n such expressions;

$$\frac{1}{\sigma^n(2\pi)^{n/2}} \exp\ \frac{[-\sum(x_i-u_0)^2]}{2\sigma^2}$$

Therefore, the likelihood ratio is;

$$\lambda = \frac{\exp[-\sum(x_i-u_0)^2/2\sigma^2]}{\exp[-\sum(x_i-u_1)^2/2\sigma^2]}$$

Taking the natural log and rearranging;

$$\sum x_i = T = [\sigma^2/(n\lambda)]/(u_1+u_0) + n[(u_1+u_0)/2]$$

In general form $T = h_i + ns$, $i = 0,1$. Decisions are based on this function by substituting the respective values for λ_0 and λ_1. Onsager (1976) and Waters (1955) provide formulas for calculating $f(\lambda)$ for binomial, Poison and negative binomial distributions.

B. Operating Characteristic Function and the Average Sample Number Function

Two other calculations of importance are the operating characteristic function (OC) and the average sample number function (ASN). Once the levels α and β have been set for an SPRT, the probability that the test will terminate with the acceptance of the null hypothesis depends on the distribution of the sampled variable. If the probability density function of x is $f(x,\theta)$ then the probability of accepting H_0 is a function of θ. The OC curve therefore gives the level of probability of accepting H_0 or H_1 for any true parameter of the actual population.

The following argument used for the derivation of OC is due to

Baker (1950). A normal distribution is assumed. Suppose a sequential test is based on the quantities u_0, u_1, α, β and σ. We wish to know α^1 for accepting H_1 when the true mean is u. When a large number of sequential tests are performed the proportion α^1 of tests terminating with acceptance of H_1 is dependent on h_0, h_1, and s. An OC curve based on u_0^1, u_1^1, α^1 and β^1 which gives the same value of h_0, h_1 and s would therefore be the same. Suppose $u_0^1 = u$ and recall that;

$$s = (u_0 + u_1)/2$$

We can see that u_1^1 must equal $u_0 + u_1 - u$ for this to hold. By the same reasoning in order for h_0 and h_1 to remain unchanged and remembering $\lambda_0 = (1-\alpha)/\beta$ and $\lambda_1 = \alpha/(1-\beta)$;

(1) $h_0 = [-\sigma^2/(u_1+u_0-2u)]\ln[(1-\alpha^1)/\beta^1] = [-\sigma^2/(u_1-u_0)]\ln[(1-\alpha)/\beta]$

(2) $h_1 = [-\sigma^2/(u_1+u_0-2u)]\ln[(1-\beta^1)/\alpha^1] = [-\sigma^2/(u_1-u_0)]\ln[(1-\beta)/\alpha]$

Write $A = (1-\beta)/\alpha$; $B = (1-\alpha)/\beta$; $t = (u_1+u_0-2u)/(u_1+u_0)$. Then (1) and (2) may be written as;

$$(1-\alpha^1)/\beta^1 = B^t \text{ and } (1-\beta^1)/\alpha^1 = A^t$$

Solving for α^1;

$$\alpha^1 = (B^t-1)/(A^tB^t-1)$$

Thus, values of t are calculated for desired values of u and α^1 calculated accordingly.

The average number of observations (N) required by a sequential test is dependent on the distribution of x which is determined by the parameter θ. The expected value of (N) therefore depends on θ. Davis (1958) provides the following derivation.

We have shown that the SPRT is a decision based on the total number of observations exceeding linear limits;

$$h_0 + ns$$

$h_1 + ns$

If we reduce the test to a scoring procedure taking s as the origin of all observations and calling resultant numbers the score, the test terminates when the scores reach either of the limits h_0 or h_1. If α is the probability of accepting H_1 when $\theta = \theta^1$, on the average $Q(1-\alpha)$ tests will terminate with a score h_0 and $Q\alpha$ tests will terminate with score h_1 where Q is the total tests performed. The total of all scores for all observations NT is $Q(1-\alpha)h_0 + Q\alpha h_1$. In practice, this is only approximate as the scores will normally exceed these bounds. The total score for NT observations is NT times the average score. Therefore the average sample number when $\theta = \theta^1$ is;

$$\frac{NT}{Q} \quad \underbrace{(1-\alpha)h_0 + \alpha h_1}_{\text{average score per observation when } \theta = \theta^1} \quad \frac{(1-\alpha)h_0 + \alpha h_1}{\theta^1 - s}$$

In summary, the OC function describes how well the test procedure achieves its objective of making correct decisions and the ASN function represents the price paid in terms of the number of samples required for the test. Onsager (1976) and Waters (1955) provide formulas for easy calculation of OC and ASN.

C. Application Using the Negative Binomial

Use of the SPRT is dependent on knowing the distribution of the population to be sampled and that for the characteristics of a specific SPRT to remain stable in time and space so must the distribution. Various mathematical distributions have been used to describe biological patterns. Mathematically, a distribution function describes a particular random variable. As few plant and animal populations are randomly distributed the negative binomial (NBD) has been used extensively to describe the

the distribution of aggregated populations (Iwao and Kuno 1971). Much of this discussion will center about it.

Individual terms of the expected frequency distribution of NBD are given by n(P(x)) where n is the sample size and P(x) is the probability of accurance of the random variable x. Individual terms of function P(x) are given by;

$$\frac{k!q^{k-x}p^{x}}{x!(k-x)!}$$

where q = 1-p, k is a measure of dispersion and $p = \bar{x}/k$. Clearly for a given $\bar{x}$, the distribution is dependent on k and concurrently so are the characteristics of a SPRT based on this distribution.

Inaccurrate estimation of k has obvious implication in calculating SPRT decision boundaries. Knight (1967) illustrated the divergence of these boundaries with different values of k for similar population means. Less obvious is the impact of divergence of k in the population being sampled from the estimate of k used to calculate the decision boundaries on the power of a specific test.

That k, or aggregation tendancy of a species, may not be independent of the density, quality or age of the population or the environment in which it exists is well documented (Abrogast and Mullen 1978, Berthet and Gerard 1965, Breyer 1968, Harcourt 1961, 1963, 1964, Iwao and Kuno 1971, Kobayashi 1968, Kuno 1963, Shiyomi 1976, Wadley 1950). Bliss and Owen (1958) derived a technique for obtaining from a series of $\hat{k}$, the most likely $\hat{k}$ which they called common $\hat{k}$. However, this does not eliminate the potential problems which variable k values may incur on the accurracy of the SPRT. Stevens et al (1976) recently concluded that the common $\hat{k}$ may not be an adequate substitute for the parameter k in all

the different k that describe a population. In addition, they concluded that sequential sampling plans dependent on a common $\hat{k}$ should not be applied at times when the common $\hat{k}$ cannot be used as an adequate substitute for k. However, emperical evidence was not provided.

While this discussion deals primarily with NBD the general implications need not be so restricted. Distribution patterns may completely change with changes in population density (see for example Ellenbeger and Cameron 1978) and the resultant affect on an SPRT based on the original distribution will be severe.

Barnard's Sequential Procedure

Barnard's sequential procedure is similar to the SPRT in that a likelihood ratio is compared to upper and lower bounds and decisions are made accordingly. However, it is assumed that x is distributed normally with unspecified variance and mean $u = \delta\sigma$ where δ is the non-centrality parameter;

$$\delta = (u - u_0)/\sigma$$

The hypotheses are stated as;

H_0: $u = \delta\sigma$

H_1: $u = \delta^1\sigma$

with δ and δ^1 specified. That this is logical can be illustrated as follows:

If we wish to determine the probability that x is less than some value z and x is normally distributed with mean u and standard deviation σ, the probability that x is less than z, $P(x<z)$, equals $f_n[(z-u)/\sigma]$ where $f_n(x)$ is the cumulative distribution for the standard normal distribution. By subtracting $\bar{x}$ from all x our interest is transformed to $P(x<0)$ which equals $f_n(\delta)$. Thus, the question we are asking is whether

$(u - u_0)/\sigma = \delta$ or $(u - u_0)/\sigma = \delta^1$.

With n observations, x_i, i = 1,n the likelihood function L is;

$$\prod_{i=1}^{n} (1/(2\pi\sigma)^{\frac{1}{2}})\exp((x_i-u)^2/2\sigma^2)$$

Which may be stated as;

$$L(x.s/u, \sigma)$$

because x. and s are jointly sufficient statistics for u and σ. Considering the ratio $t = x.(n/s)^{\frac{1}{2}}$, on H_0 t has a non-central t distribution with (n-1) degrees of freedom and parameter δ. On H_1 the non-centrality parameter is δ^1. With the probability density indicated by f the likelihood ratio criteria, considering only the distribution of t is;

$$\lambda(t/\delta, \delta^1) = f(t/\delta^1, n)/f(t/\delta, n)$$

Evaluating $\lambda(t/\delta, \delta^1)$ as a function of t for each value of n requires a difficult series of approximations. Tables are available for comparing the test statistic $U = \sum(x-u_0)/\sum(x-u_0)^2$ for various values of α, β and difference in δ which it is important to detect.

Little is known about the OC and ASN for these tests. The test is not linear and therefore usual ASN formulas are not applicable. Rushton (1950) provides means for obtaining the lower bound to the mean sample size.

Fowler's Sequential Procedure

Fowler and O'Regan (1974) present a truncated sequential t-test derived by employing Monte Carlo procedures to approximate the distribution of the conditional test statistic at each stage of the test. A truncated test ensures that a decision will be reached before or at a specified number of observations. Their test is constructed by specifying α, the truncation point n_0 and a probability boundary pattern. This boundary pattern establishes the probability of accepting and rejecting H_0 (H_0

being true) at each stage of the test such that the overall probability of rejecting H_0 with H_0 true is α.

The test statistic is defined as;

$$d_n = (\bar{x} - u_0) / \sum(x_i - \bar{x})^2/n(n-1)$$

which has a t distribution with one degree of freedom for d_1 and an unknown conditional distribution for d_n $(n>1)$. Decision points for rejection and acceptance of H_0 based on d_n were approximated using Monte Carlo procedures. Null and alternate hypotheses are stated in terms of δ;

$$H_0: \quad u = u_0 \text{ or } \delta = \delta_0 = 0$$

$$H_1: \quad u = u_1 \text{ or } \delta = \delta_1, \ (\delta_1 > 0)$$

OC and ASN functions of each test are approximated with simulation techniques.

The sequential t procedures assume that the population being sampled is approximately normal. Divergence from this assumption is known to alter the power of a test employing this statistic (Pearson and Please 1975).

Iwao's Sequential Procedure

Iwao (1968) proposed the regression of Lloyd's mean crowding on mean density as a method for analyzing aggregation patterns. Lloyd (1967) established the parameter mean crowding as the mean number per individual of other individuals in the same quadrat. The parameter is defined as;

$$\overset{*}{u} = \sum_{j=1}^{Q} x_j (x_j - 1) \sum_{j=1}^{Q} x_j$$

where Q is the total number of quadrats and x_j is the number of individuals in the j^{th} quadrat. Mean density is related to it through;

$$\overset{*}{u} = u + (\sigma^2/(u-1))$$

The parameter is estimated by substituting the sample mean and variance $\bar{x}$, s^2 for u, σ^2. The relationship;

$$\overset{*}{u} = A + Bu$$

has been shown to be linear in a wide variety of applications (Iwao 1968, Iwao and Kuno 1971). If a population distribution follows a Poisson series the regression line passes through the origin (A=0) and its slope B is equal to unity. For a negative binomial with a common k, A = 0 and B = 1 + (1/k). Underdispursed or completely uniform distributions are characterized by $\overset{*}{u}$ taking zero up to u = 1 and increasing along the linear regression of slope B ≠ 1. Through extrapolation of the regression line from u ≥ 1, A is -1.

The value of the intercept A may be interpreted as the number of organisms which would live together with A other individuals at some infinitessimally small density. Iwao (1968) categorized this as the "Index of Basic Contagion". The slope B is an index of the spatial pattern of habitat use by individuals or groups of individuals in relation to their density. This is called the "Density - Contagiousness Coefficient". Both indices are necessary to describe a distribution. In relation to the problem currently under investigation, a linear relationship has been shown to hold even when k varies with population density in a negative binomial (Iwao and Kuno 1971).

As previously indicated, mean crowding, $\overset{*}{u}$, is related to population variance (σ^2) by;

$$\overset{*}{u} = u + (\sigma^2/(u-1))$$

Therefore;

$$\sigma^2 = u(\overset{*}{u}-u+1)$$

Substituting the mean crowding to mean density relationship;

$$\sigma^2 = (A+1)u + (B-1)u^2$$

The half width of a confidence interval is given by $d = ts_{\bar{x}}$ where t is the value of the normal deviate corresponding to a desired confidence probability (Student's t). Employing the variance relationship Iwao and Kuno (1968) calculated d as;

$$d = t(((A+1)u + (B-1)u^2/n)^{\frac{1}{2}})$$

where n is the number of samples.

As illustrated by Iwao (1975) this relationship may be used to calculate a sequential sampling procedure based on whether observations fall within desired confidence limits based on a hypothesized mean (u_0). The upper limit is defined as;

$$T_u = nu_0 + t(n((A+1)u_0 + (B-1)u_0^2)^{\frac{1}{2}})$$

and the lower limit as;

$$T_1 = nu_0 - t(n((A+1)u_0 + (B-1)u_0^2)^{\frac{1}{2}})$$

Decisions are made with the assigned confidence limit that $\bar{x} > u_0$ when $T > T_u$ and $\bar{x} < u_0$ when $T < T_1$ where T is the total number of individuals.

As in other sequential procedures when $\bar{x}$ is very close to u_0 a large number of samples may be required for termination. However, the maximum number of observations required for a desired confidence limit, d, is;

$$n_{max} = (t^2/d^2)((A+1)u_0 + (B-1)u_0^2)$$

This method is intuitively appealing because of the reported stability of the mean crowding - mean density relationship. This clearly circumvents problems encountered due to variable distribution parameters and deviations from the normality assumption.

METHODS

We are now in a position to investigate the properties of the aforementioned tests. Specificly; 1) the power (OC) defined as the probability

of correctly choosing between two hypotheses about a true population parameter and average number of samples (ASN) required for each test and, 2) the effect of deviations from assumptions inherent to tests are investigated. In brief, the experimental design was to sample a population and then use this sample as a population with which to simulate sampling. The sampled population consisted of FTC egg masses located on 76 cm (30 inch) branch samples. All possible samples were taken from the upper crown of sugar maple (Acer saccharum) trees by felling the trees and then carefully pruning off the branches. The mean number of branch samples per tree was 148.7 ($s = 41.3$). Data collection took place in August and September 1978 in an area surrounding Pellston, Michigan. In total, nineteen complete trees were enumerated.

Mean crowding and mean density were calculated for each tree and a regression of these parameters formulated to describe the egg mass distribution.

For the simulation, six routines were developed. These are described below. A listing of each routine may be found in respective appendices.

1) STAT calculates statistics; mean, variance and estimate of k through the maximum likelihood method of the population (Appendix pg. 39).

2) NEGBIN distributes egg masses among 500 samples according to a negative binomial. The mean and k are specified exogenous from the routine (Appendix pg. 42).

3) SEQUAN randomly samples the population 1000 times and makes decisions regarding the mean with SPRT. The null hypothesis is given as $\bar{x} \leq 0.2$ egg masses per branch and the alternate hypothesis as $\bar{x} \geq 0.5$. The latter density corresponds to the reported level at which heavy

defoliation by FTC on sugar maple can be expected at the start of an infestation (Connola et al 1957, Marshall and Hoffard 1976). The probability of Type I and II errors are 0.1. Decision boundaries and nominal values of the OC and ASN based on all data points are given in appendix figure 1. Because the procedure is open ended, during the simulation sampling was arbitrarily truncated at 50 observations. If no decision had been made after 50 observations, output indicated such (Appendix pg. 45).

4) BTTEST randomly samples the population 1000 times and makes decisions regarding the mean with BTTEST. In order to facilitate use of tables, error levels were set at 0.05. The null and alternate hypotheses are given as $\bar{x} = 0.5$ and $\bar{x} > 0.5$ egg masses per branch respectively. In addition, the difference, stated in terms of standard deviations D, ($\delta = D\sigma$), which it is important to detect, was set at 0.5. Decision boundaries are given in appendix figure 2. As in SEQUAN sampling is truncated at 50 observations (Appendix pg. 47).

5) TTEST randomly samples the population 1000 times and makes decisions regarding the mean with FTTEST. The null and alternate hypotheses and error levels are identical to those in BTTEST. Because this is a truncated test a decision is assured at the termination point of 10 observations. Decision boundaries are given in appendix figure 3. The probability boundary pattern employed specifies α at each decision stage to increase at a constantly increasing rate (Appendix pg. 50)

6) SEQRT randomly samples the population 1000 times and makes decisions regarding the mean with ITEST. The null and alternate hypotheses are $\bar{x} < 0.5$ and $\bar{x} > 0.5$ egg masses per branch respectively. Type I and II errors are set at 0.1. Decision boundaries are illustrated in appendix figure 4. Sampling was again terminated at 50 observations

(Appendix pg. 53).

Three simulations were executed using the described routines: 1) A comparison of the OC and ASN of the four sequential procedures assuming the population distribution was constant regardless of population density; 2) the effect of distribution varying with density on OC of SPRT; and 3) a comparison of sample unit sizes as applied to FTTEST.

RESULTS

Mean Crowding and Mean Density and Comparison of OC and ASN

Regression of mean crowding on mean density (fig. 1) provided the relationship;

$$\overset{*}{\bar{x}} = -0.0052 + 1.695(\bar{x})$$

While NBD is characterized with A = 0, A of -0.0052 in this instance meets this requirement considering the variability about the regression line. Therefore, the data is described by NBD with $\hat{k} = 1.449$. From the original sample a subsample of 500 with $\bar{x} = 0.222$, $s^2 = .273$ and $\hat{k}$ of 1.0768 was randomly selected for use in the simulation. No loss in accuracy of the results is incured by using a population of n = 500, however, computations and costs are substantially reduced.

Simulation results are presented in figures 2, 3 and 4. OC curves were adjusted for no decision cases. Only SPRT and ITEST possessed useful OC curves (fig. 2). That BTTEST and FTTEST were inferior in this regard is not surprising considering the skewed distribution from which samples were drawn.

A complete picture of the usefulness of a sequential procedure must also include the ASN function. In the simulation this is comprised of two components; the average number of samples per decision (fig. 3) and the proportion of no decisions after the truncation point of 50

samples (fig. 4). Although the sequential procedure may not have reached a decision stage prior to the truncation point, gathered information is still usefull. Both SPRT and ITEST exhibited similar variations about an estimated mean, however, the number of times this estimate is used differed between tests. At the critical density the 95 percent confidence interval half width is approximately 0.2. Assuming a constant distribution this is calculated by knowing the actual population variance at a density of 0.5 and calculating the standard deviation of the mean with a sample size of 50. An estimate with sample size of 50 is therefore 0.5 ± 0.2 egg masses. SPRT must be considered superior in this regard as this estimate is used less frequently than with ITEST.

Effects of Changing Population Distribution

As previously stated the assumption of invariant population distribution in relation to population density is often not upheld. The second simulation was designed to investigate the effect of divergence from this assumption on SPRT. Linear relationships between k and the population mean were established. Maximum divergence are given by;

$k = 1.077 + 10(\text{Mean})$

$k = 1.077 - 0.5(\text{Mean})$

The first resulted in the population becoming more random while the second produced a more aggregated distribution. The relationships do not imply a realistic representation and are presented merely as examples. However, this has no effect on conclusions drawn from the results. The simulation employed NEGBIN to update the population distribution with each change in mean density.

Changes in OC associated with the largest change of k in relation to the mean are presented in figure 5. A more random distribution

actually improved the power of a test considering it important to detect population levels exceeding the critical density. A tendancy toward greater aggregation in relation to the mean had an opposite effect. Differences in OC due to changes in k must be considered in light of respective changes in population distribution. From figure 6 it is evident that variable k did not induce large changes in the distribution. Therefore, changes in OC were likewise not great. Divergence in OC may be much greater for other NBD.

Effects of Different Sample Unit Sizes on FTTEST

Obtaining estimates of parameters required to describe a population distribution is often tedious. Also, these parameters are rarely invariant with density. For these reasons sequential t procedures are appealing. However, we have already demonstrated their ineffectiveness when the distribution is highly skewed. More specificly we are dealing in this case with a population containing a high percentage of zero values. This situation commonly presents itself when sampling insect populations. A solution to this problem is now presented.

Population distribution is largely a function of the size of the sample unit. By varying the sample unit, the distribution and associated mean and variance change (Elliot 1977, Pielou 1978). Employing this principle, sample surface can be varied by considering 1, 2,..., n branches as an observation. A sequential procedure may then be considered by increasing the densities defined under the null and alternate hypotheses by the same change in factor as the sample surface. Thus, if the null hypothesis is; $\bar{x} = 0.5$ with a sample surface of 1 it will be increased to $\bar{x} = 1.0$ with a sample surface of 2.

The effectiveness of the strategy was investigated with FTTEST and this comprised the third simulation. The OC and ASN functions for sample surfaces of 1 to 9 branches are given in figures 7 and 8 respectively. ASN is still expressed in terms of number of branches and not observations. Additionally, the effect of change in sample surface on distribution is portrayed in figure 9. As sample surface increased, the OC became more favorable. Though a sample surface of 9 has a superior OC to sample surface 7, the cost in terms of ASN is large. It is also evident that increasing sample surface to 9 did not normalize the distribution. Thus, FTTEST must be considered robust to some deviation from the assumption of normality.

Figure 1. Regression of mean crowding on mean density of forest tent caterpillar egg masses on 76 cm. branch samples from 19 trees. The model is: $Y = A + BX + e$ where Y = mean crowding, X = mean density, A = -0.0052, B = 1.695, $R^2 = 0.8704$.

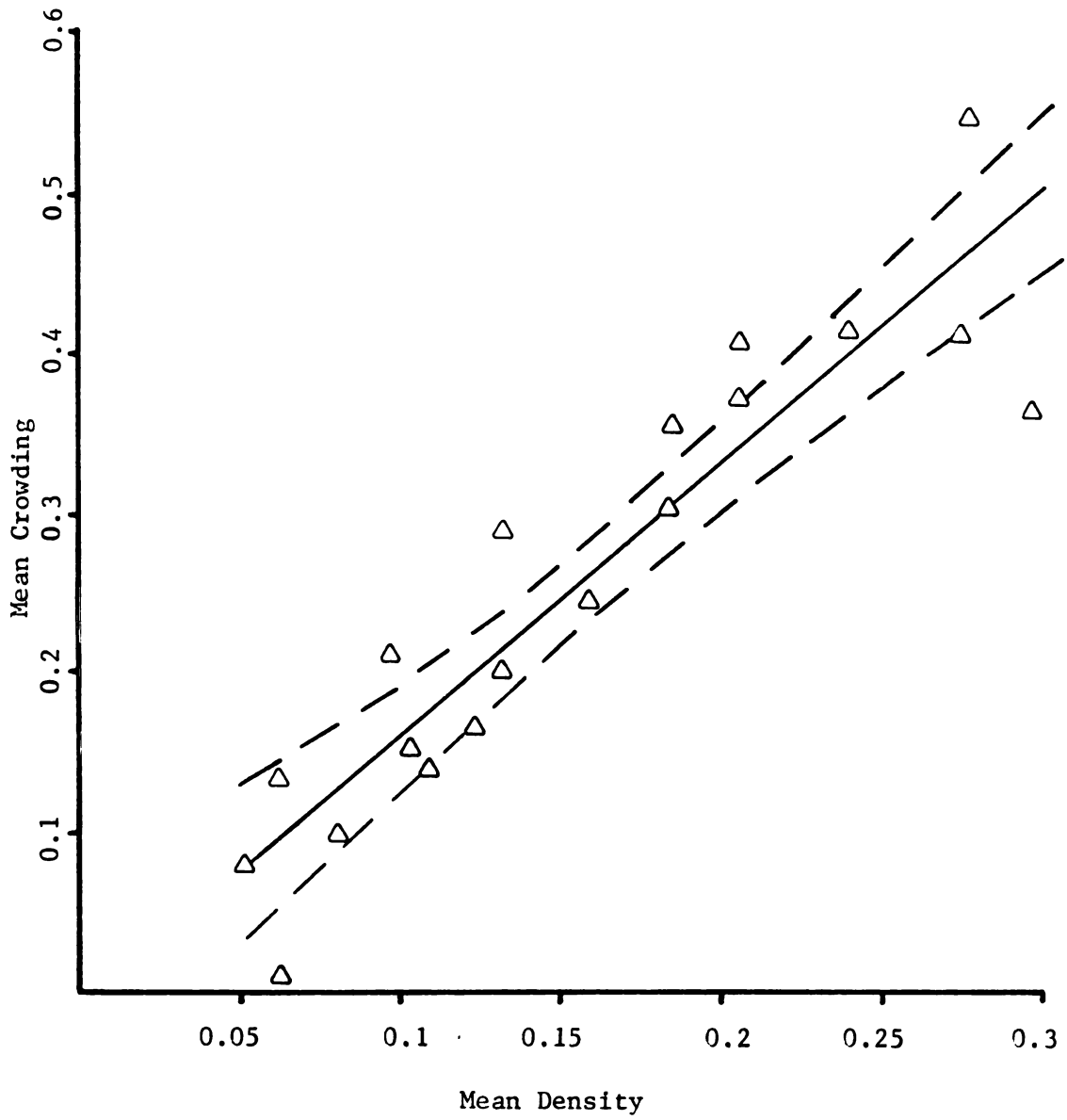


Figure 2. Adjusted OC curves for FTTEST, BTTEST, ITEST and SPRT.

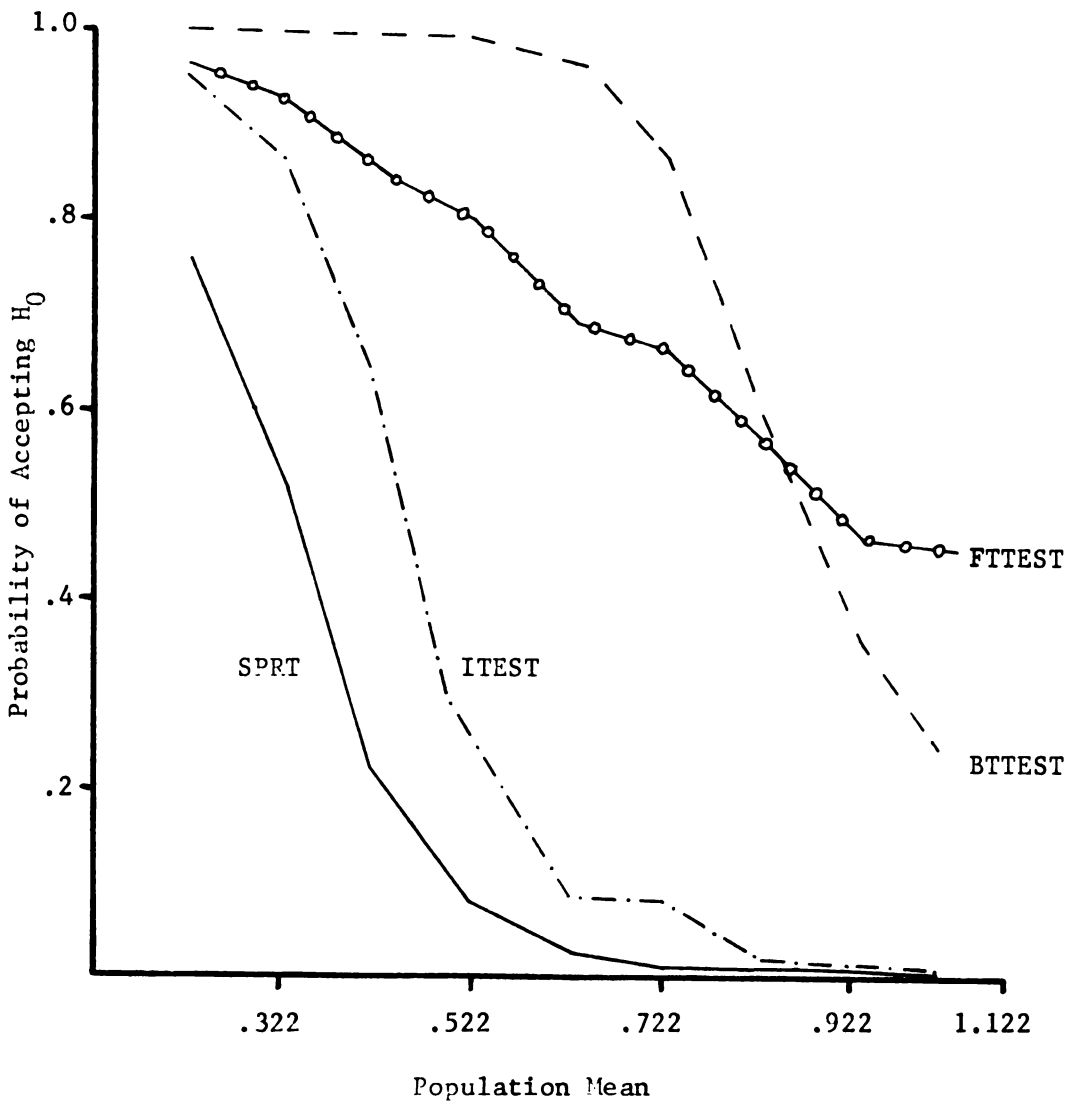


Figure 3. Average sample number in relation to population mean for FTTEST, BTTEST, ITEST and SPRT.

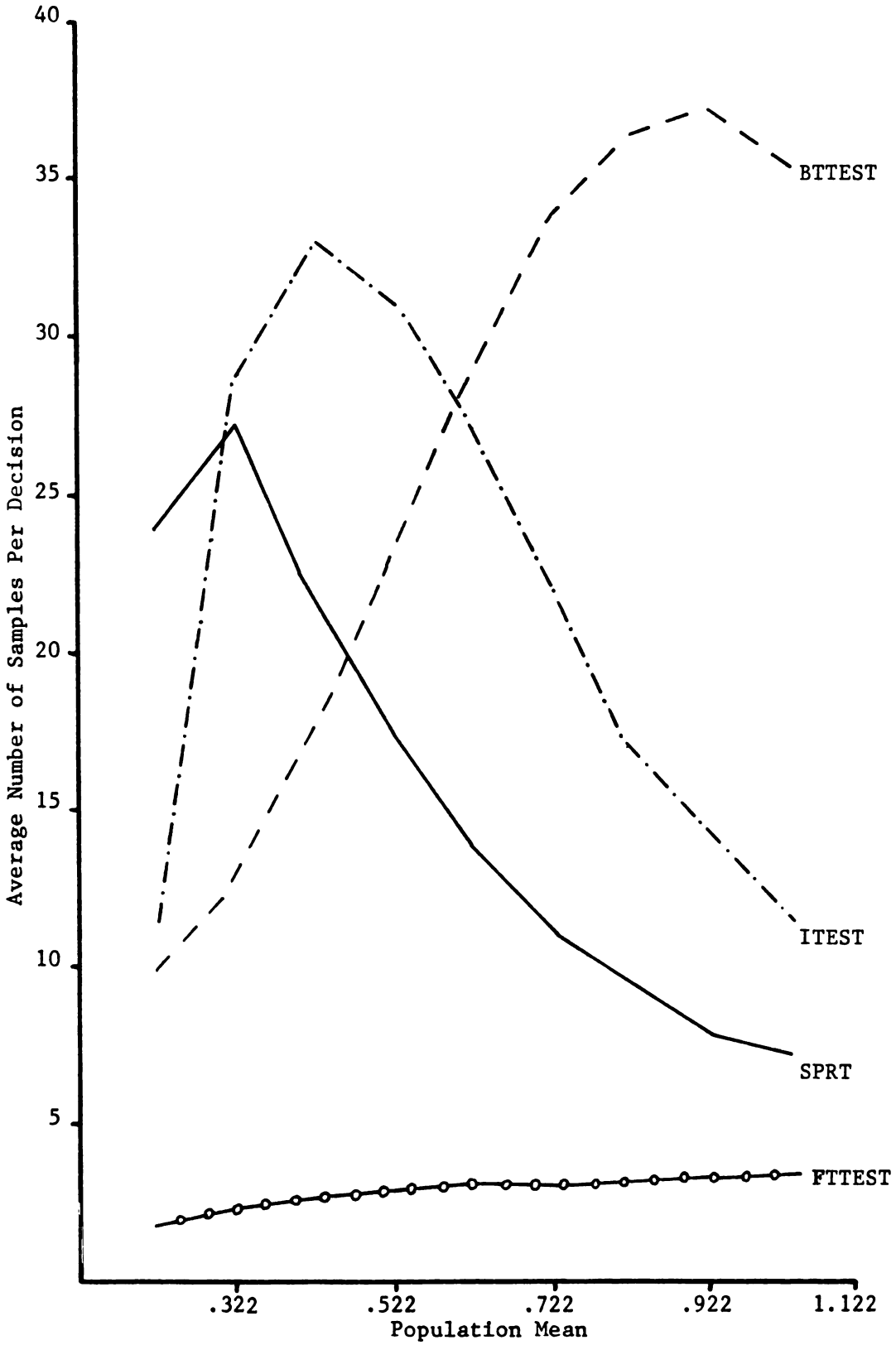


Figure 4. Proportion of no decisions after fifty observations in relation to mean density for BTTEST, ITEST and SPRT.

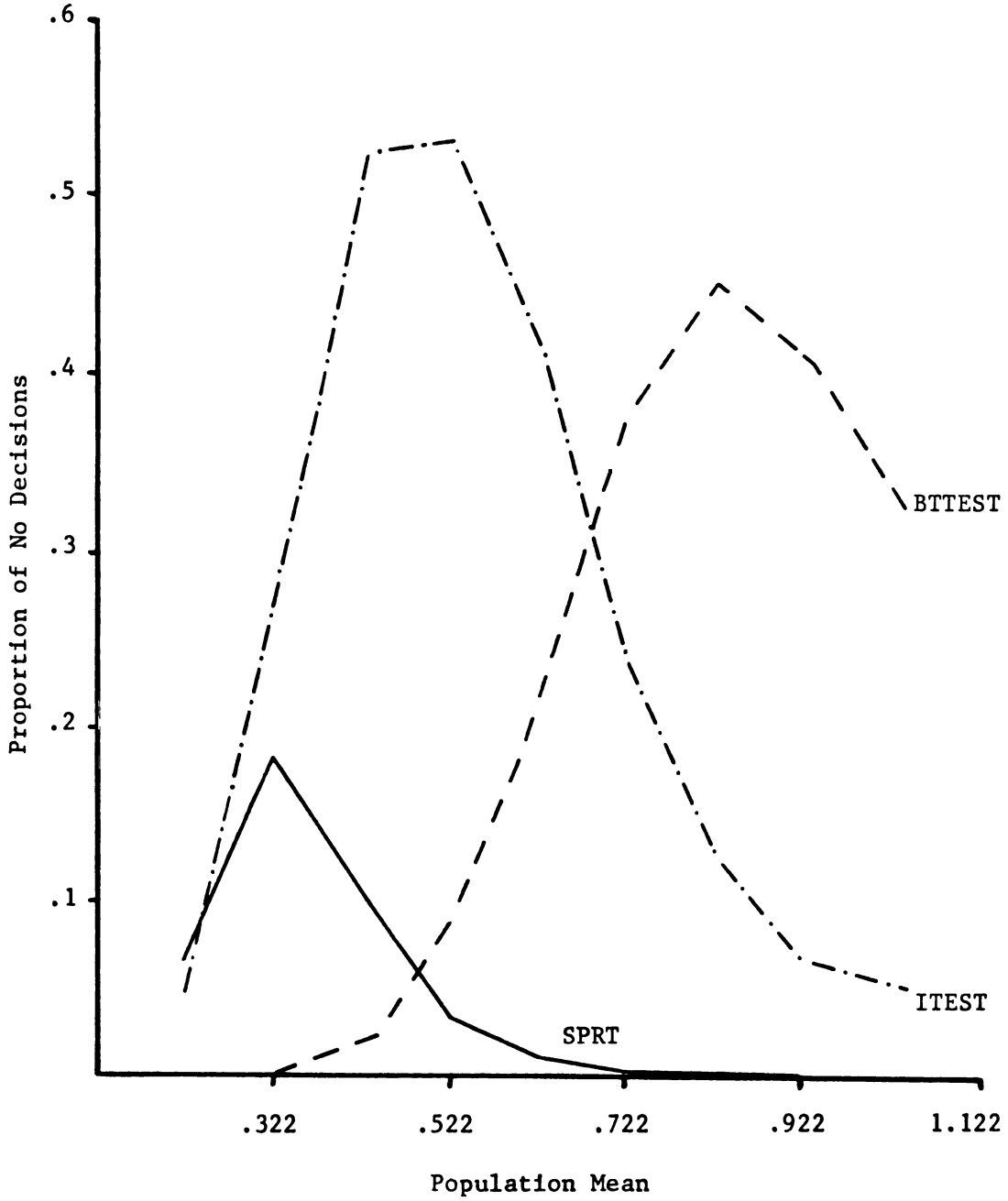


Figure 5. Probability of accepting H_0 relative to the population mean for three hypothetical mean - k relationships employing SPRT.

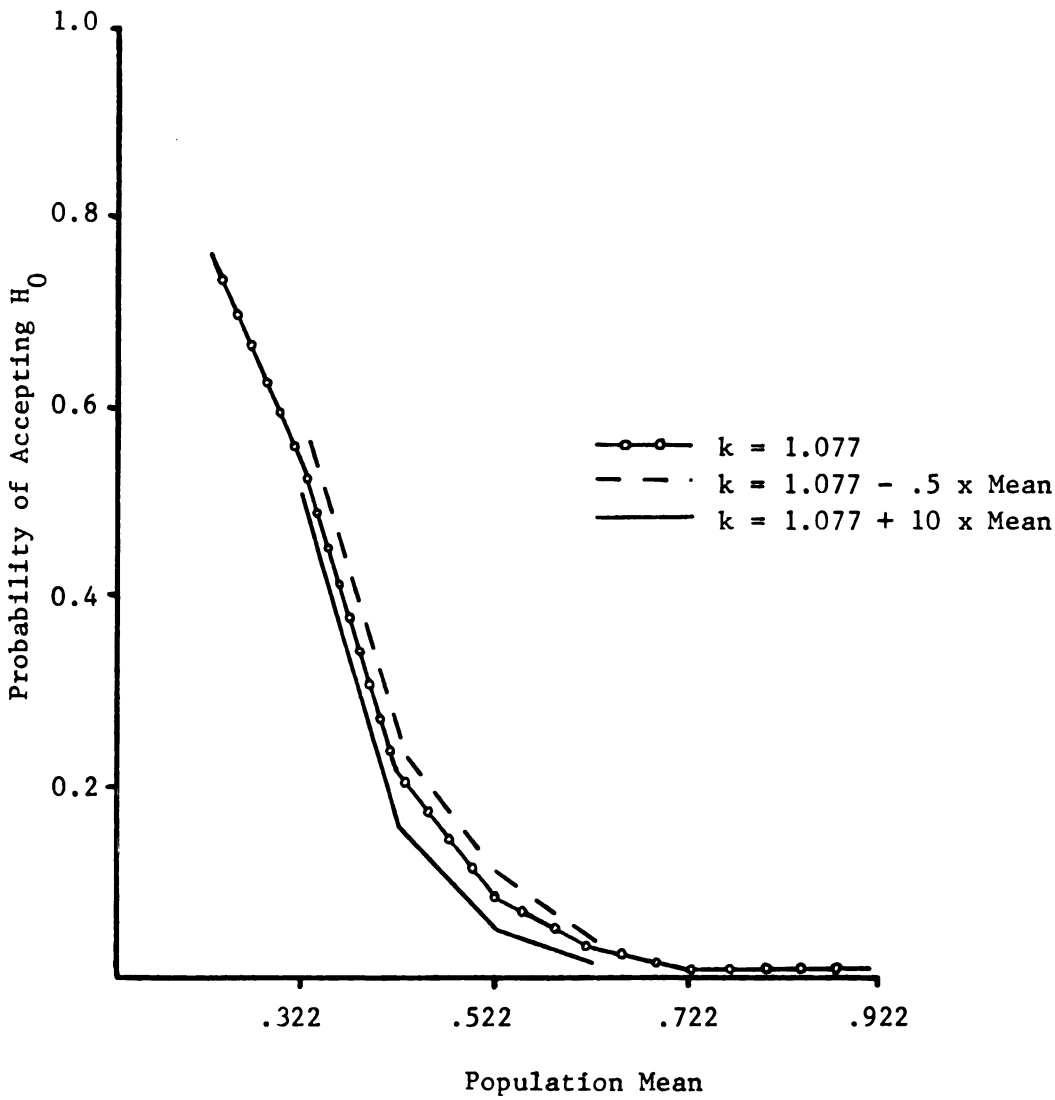


Figure 6. Effect of changes in k of negative binomial distribution on frequency distribution of forest tent caterpillar egg masses. Class refers to number of egg masses per branch sample.

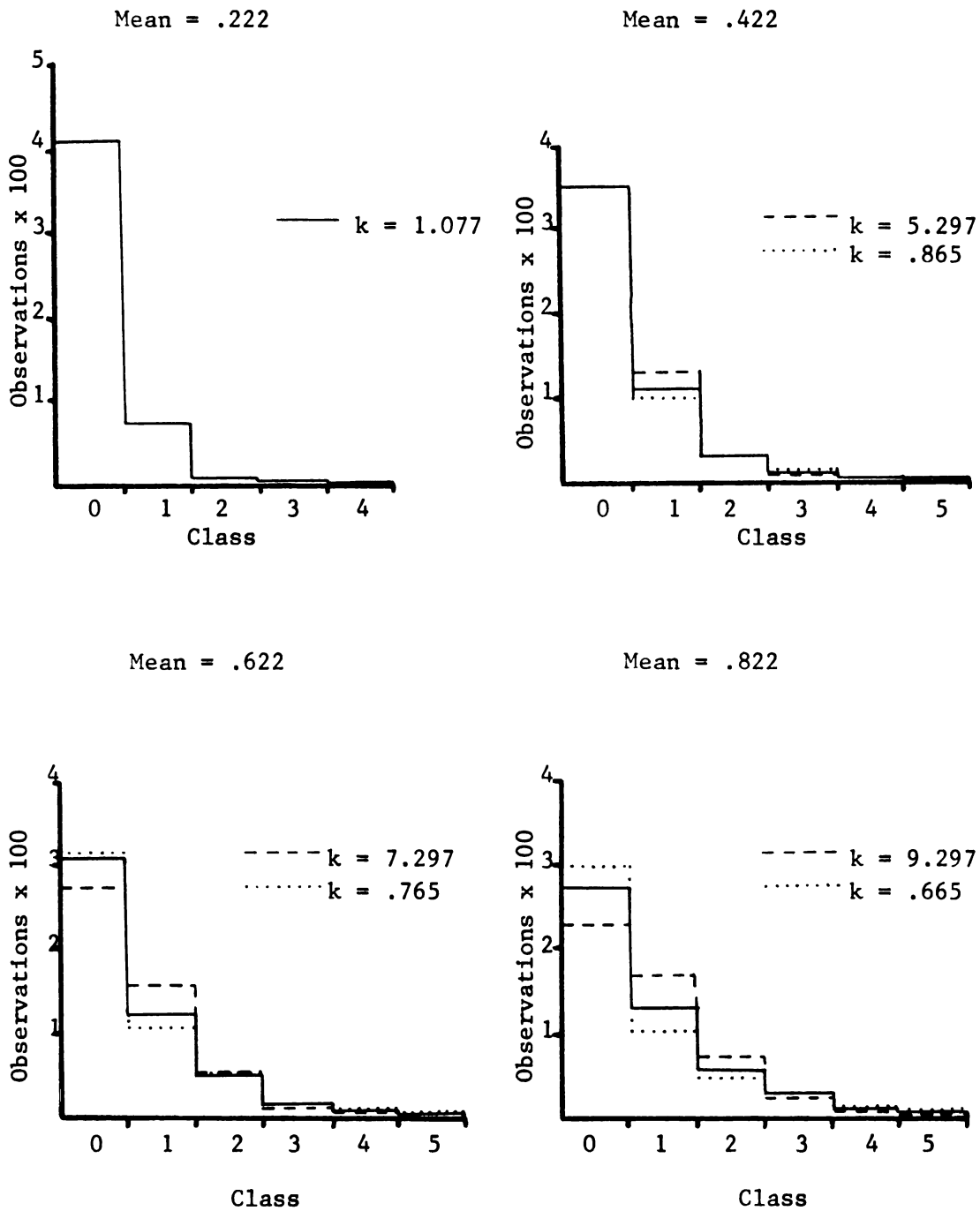
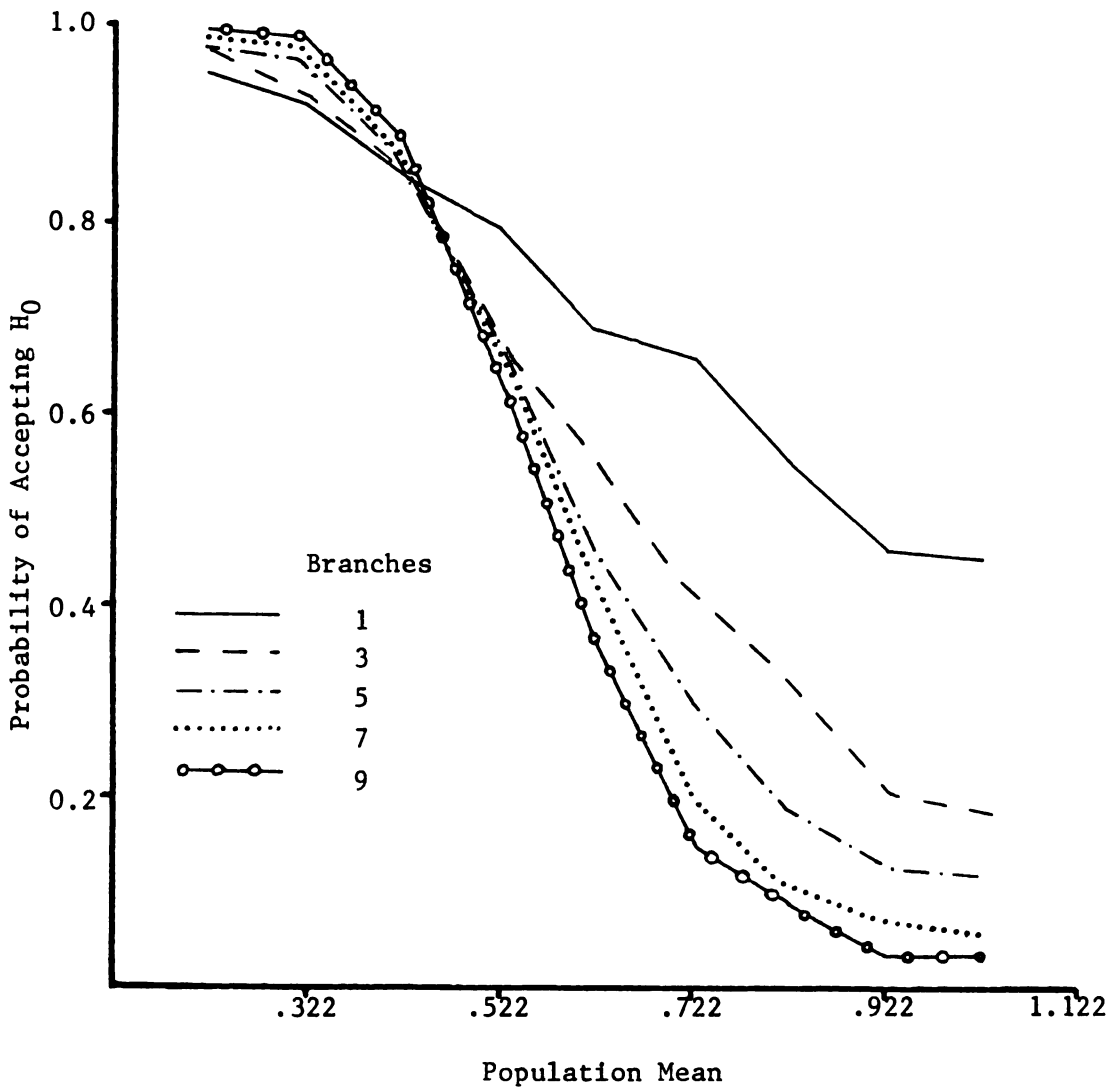


Figure 7. Probability of accepting H_0 relative to the population mean for five sample surfaces represented by branches per observation employing FTTEST.



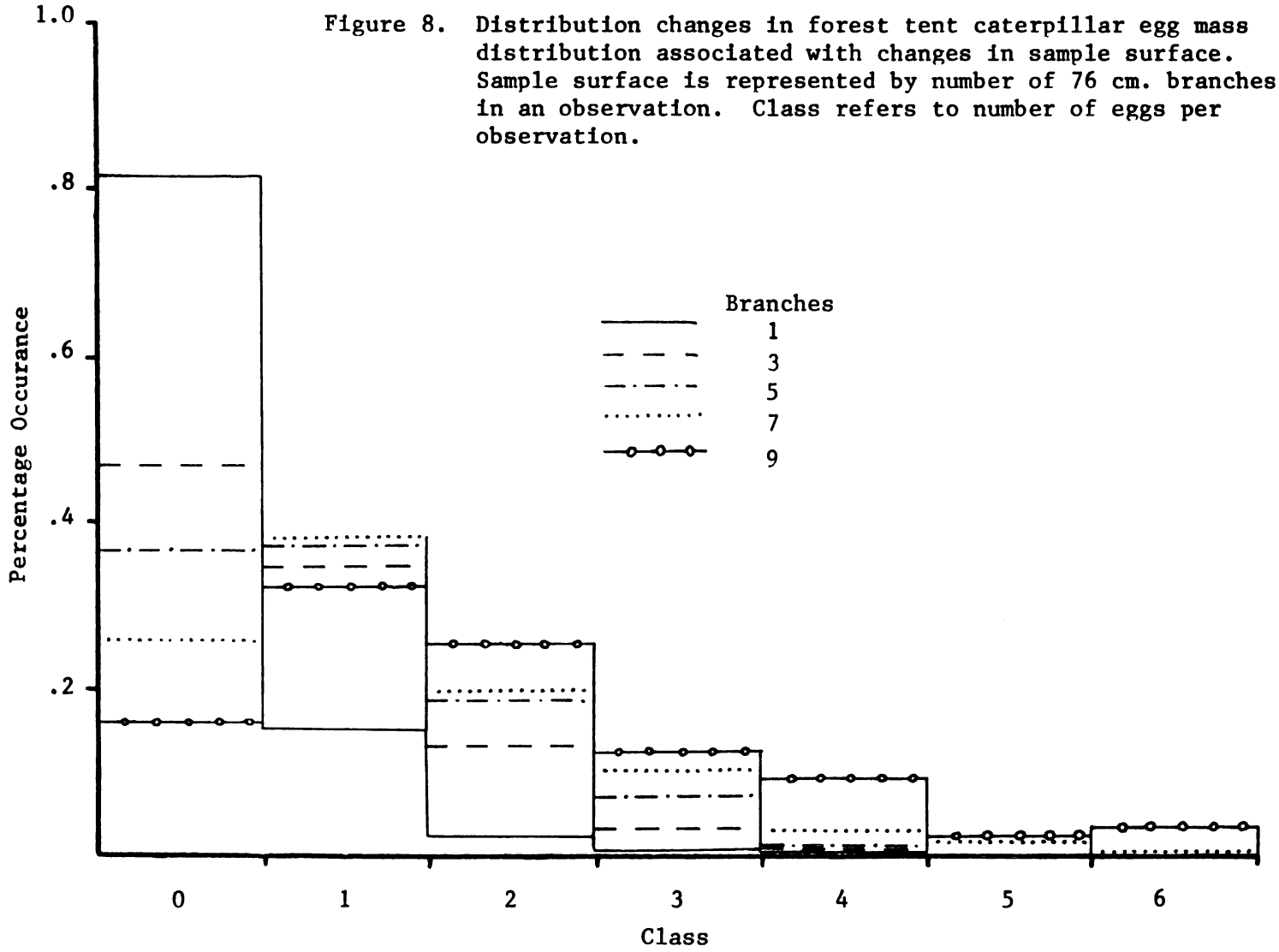
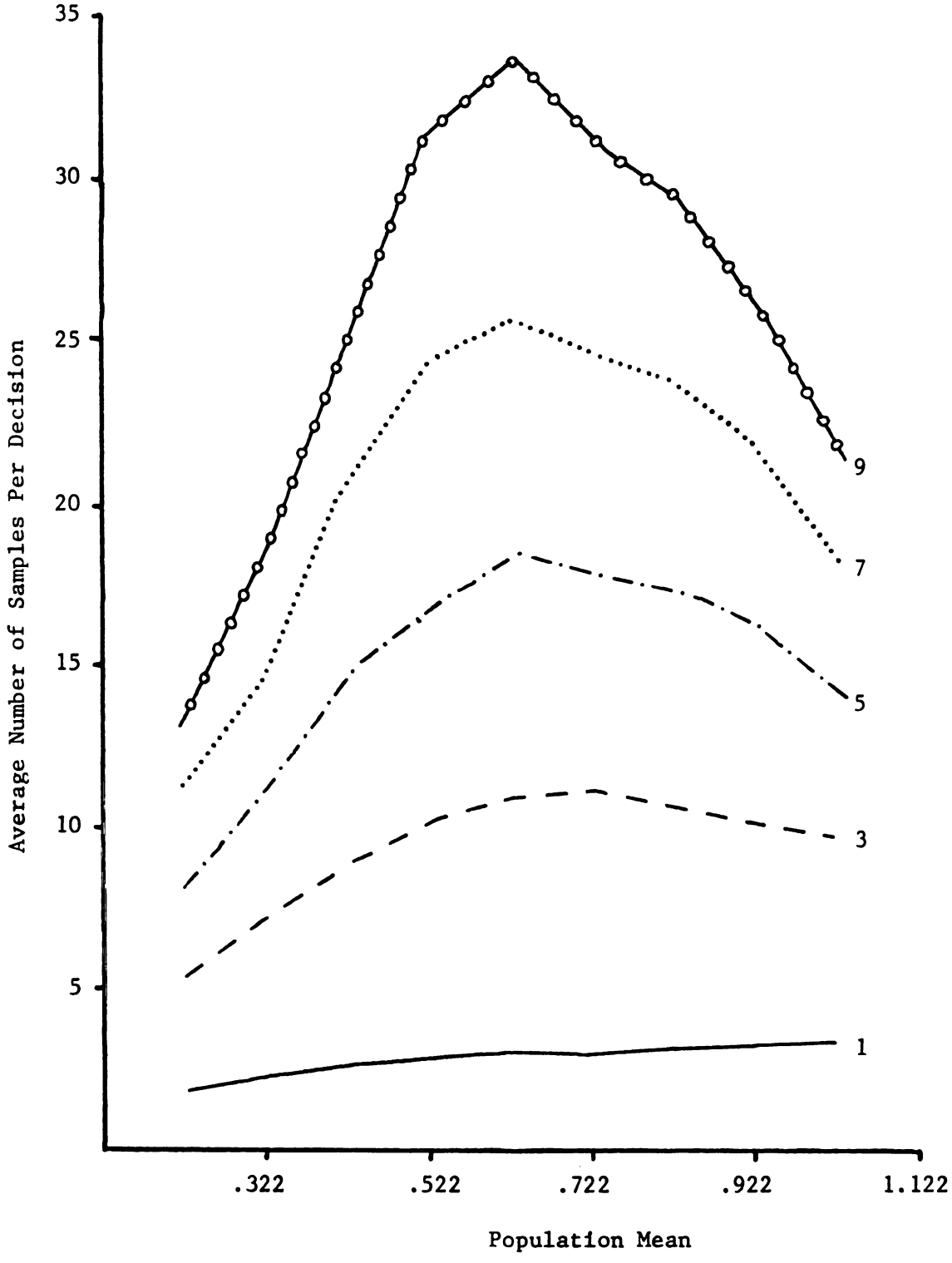


Figure 8. Distribution changes in forest tent caterpillar egg mass distribution associated with changes in sample surface. Sample surface is represented by number of 76 cm. branches in an observation. Class refers to number of eggs per observation.

Figure 9. Average sample number curves for FTTEST for five sample surfaces represented by branches per observation.



DISCUSSION AND CONCLUSIONS

Selection of a sequential procedure must be based on three criteria; the OC and ASN functions and information required about a distribution for construction of a procedure with desirable OC and ASN. The type of information required for quantifying population distributions according to a numerical function or the regression of mean crowding on mean density are similar. If an SPRT is to be employed one is limited to a negative binomial, Poisson, binomial or normal distribution as the SPRT is available only for these distributions. Additionally, it is unlikely that the parameters estimated for these distributions are invariant to population density changes. These changes in turn will affect OC and ASN of SPRT, however, as illustrated with FTC these changes may not be so severe as to render the test useless. Clearly, if a variable parameter say k of a negative binomial distribution is suspected, it behooves the investigator to have prior knowledge of how this varability will affect SPRT. By placing reasonable bounds on this variability simulation provides a logical method for obtaining this information.

Sequential procedures based on distributions described by the regression of mean crowding on mean density are appealing due to the reported linearity of the relationship over a wide range of distributions and densities. In the simulation the OC for ITEST was comparable to the OC for SPRT and to an extent superior. Error levels at low egg densities were less with this procedure though greater at high densities. The ASN was greater with ITEST, however, the truncation point may be specified to correspond with a desired confidence interval width. This width only applies to populations approximating the critical densities and the confidence interval width will change as density deviates from

this critical density.

Because the distribution relationship is defined with a linear regression, the linear fit will influence determination of confidence intervals necessary to construct the sequential test. A similar problem may occur with SPRT, however, tests for lack of fit from a mathematical distribution are better defined. Recall that decision boundaries are based on the confidence interval given by;

$$d = t(((A+1)u + (B-1)u^2)/n)^{\frac{1}{2}}$$

With stochastic variation in A and B one can be assurred that estimated d contains true d_0 by using the upper and lower confidence bounds of A and B for the upper and lower decision boundaries of the sequential test respectively. This will widen the band with which continued sampling is dictated and increase ASN. However, it also assures OC will be as specified.

Describing a distribution with a mean crowding - mean density regression also allows use of a fixed precision level sequential plan. An obvious deficiency of sequential sampling is that the procedure only classifies a population into predesignated categories. Much work requires an actual estimate of the population mean. Kuno (1968) used the variance relationship provided by the mean crowding - mean density regression to derive a sequential procedure concerned with the level of precision attained for estimating the mean. The decision boundary is given as;

$$T_n = \frac{A + 1}{D^2 - ((B-1)/n)}$$

where D equals $s_{\bar{x}}/\bar{x}$ and T_n is the total of n observations.

When the underlying distribution of a population is unknown,

sequential t procedures may be used. As demonstrated, complete ignorance of the population distribution is unacceptable. A large percentage of zero counts or a highly skewed distribution invalidates the normality assumption and destroys the test. However, with far less effort than is needed to quantify a distribution, information may be gathered to assure that the sample surface is of proper size to insure a viable OC. A comparison of OC for FTTEST and SPRT with error levels of both tests set at 0.05 is presented in figure 10. Previously α and β of SPRT were each set at 0.1 to insure reasonable ASN. FTTEST has a better OC at low densities, however, beyond the critical density of 0.5 egg masses SPRT is superior. ASN curves for both tests are compared in figures 11 and 12. Because SPRT is an open test some cases will result in no decisions and this factor must be taken into account when comparing ASN curves. In this regard, FTTEST is superior at low densities and the reverse is true at higher population levels. Finally, FTTEST will be largely invariant to changes in population distribution. Clearly, neither test is optimal in all instances. Users must weigh these attributes and base a decision for test use on criteria specific to each case.

From this study the following generalized conclusions may be drawn:

1) Attributes of SPRT are dependent on the underlying population distribution and vary with changes in this distribution. These changes may or may not be significant.

2) ITEST and FTTEST offer likely alternatives to SPRT and may be superior when confronted with a changing population distribution.

3) FTTEST requires the least information when applicable for construction.

4) Simulation provides a quick and inexpensive method for analyzing different sequential schemes.

Figure 10. Probability of accepting H_0 adjusted for no decision cases relative to the population mean for SPRT and FTTEST (α and β = 0.05, FTTEST sample surface = 7).

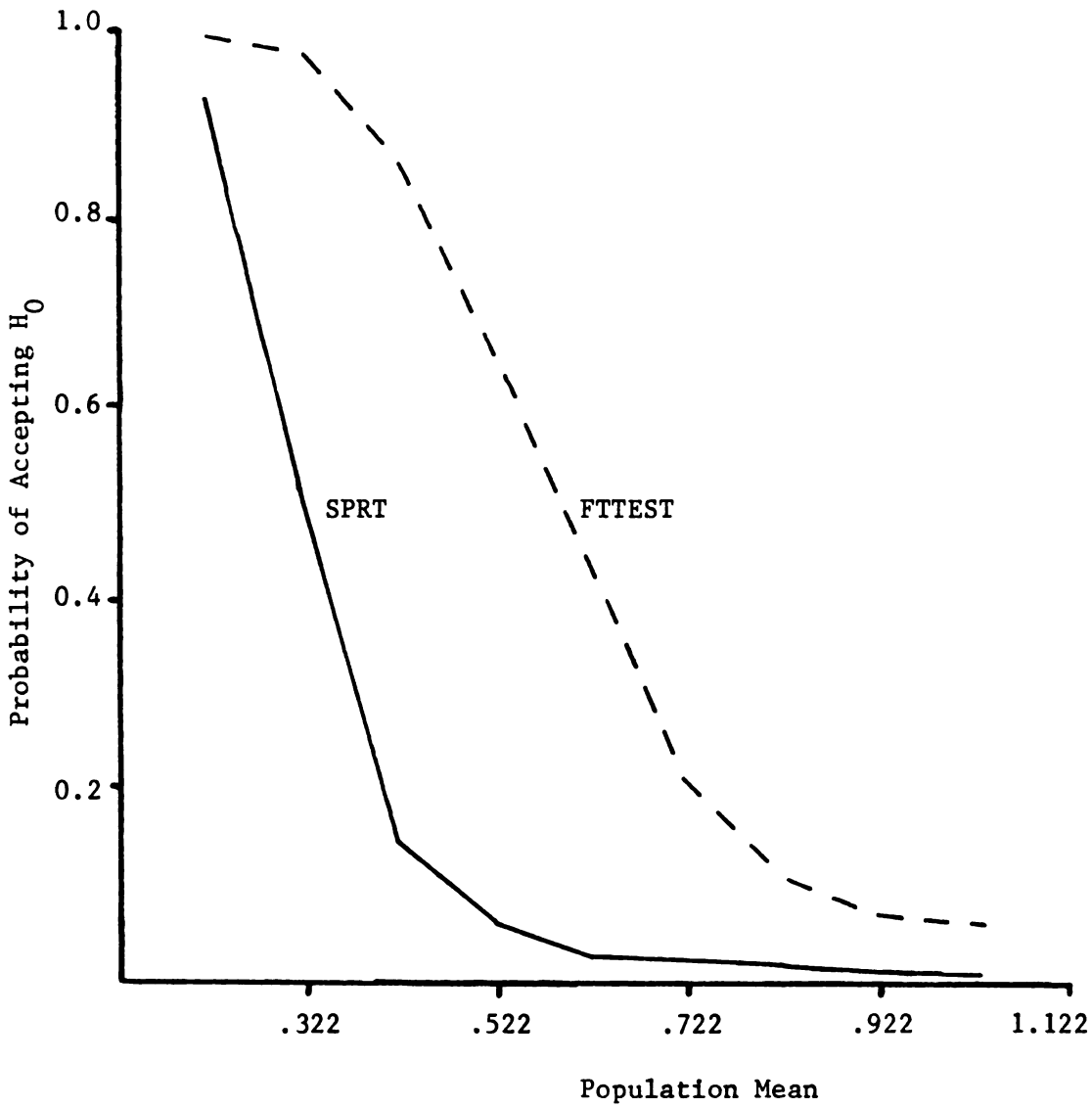


Figure 11. Average sample number in relation to population mean per fifty samples for FTTEST and SPRT (α and β = 0.05, FTTEST sample surface = 7).

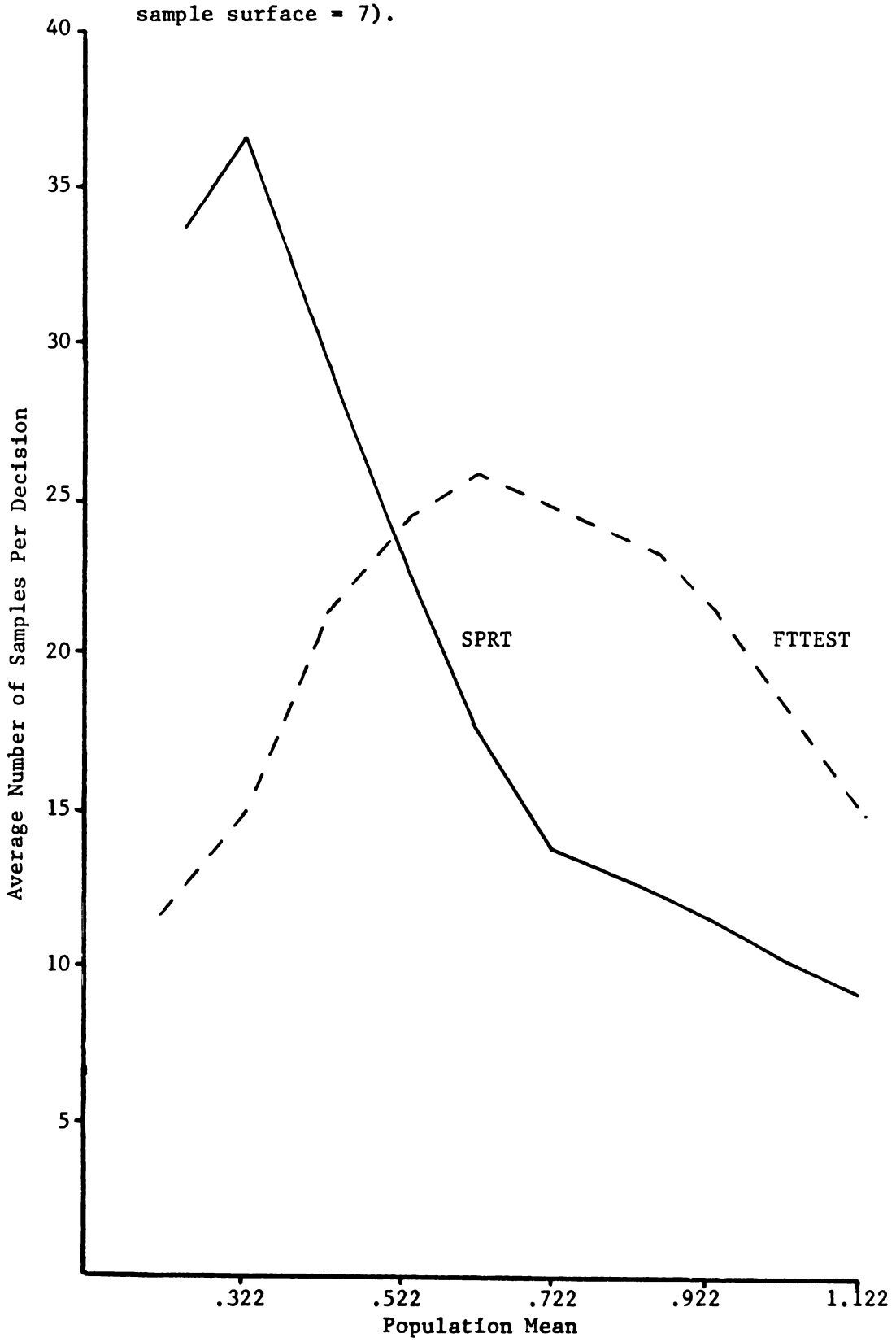
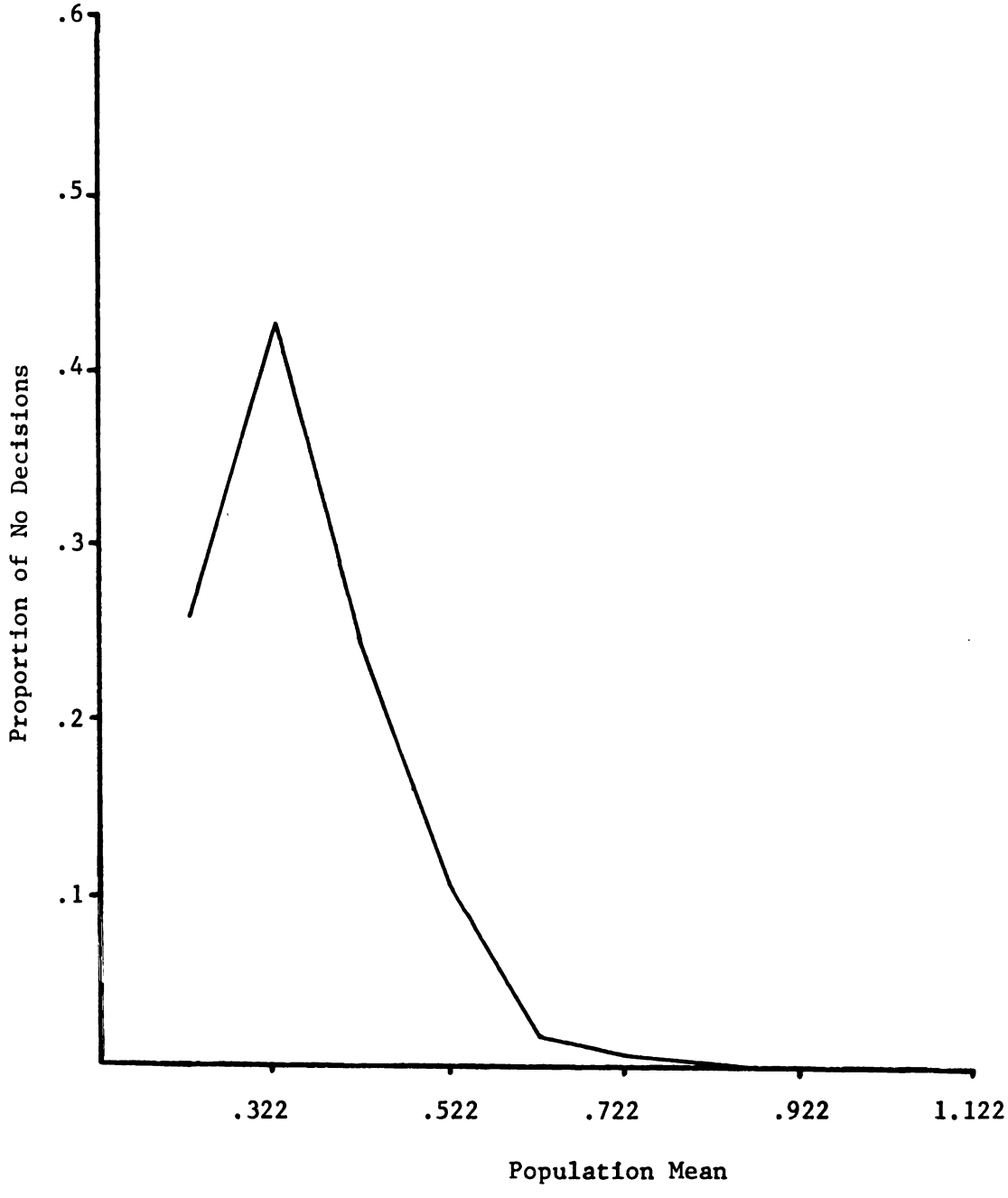
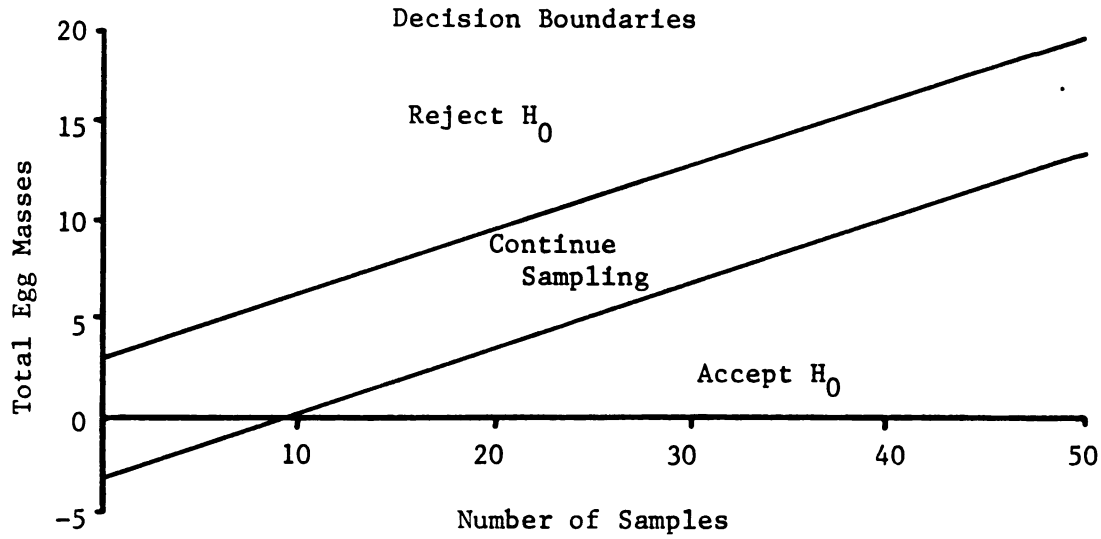


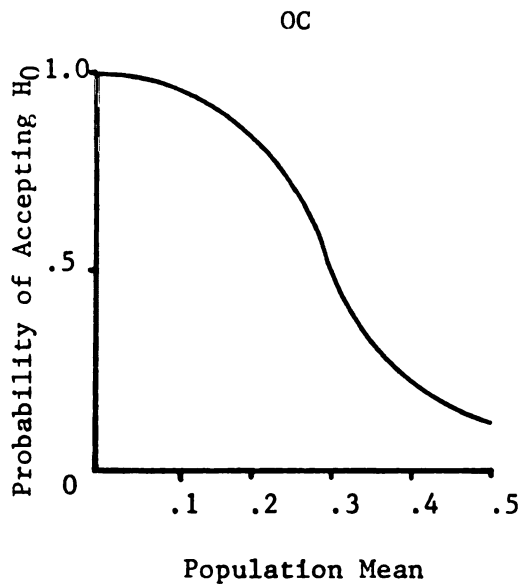
Figure 12. Proportion of no decisions after fifty observations in relation to population mean for SPRT (α and β = 0.05).



APPENDIX

Figure 1a. Decision boundaries, OC and ASN curves of SPRT for forest tent caterpillar egg mass sampling (see text for details).





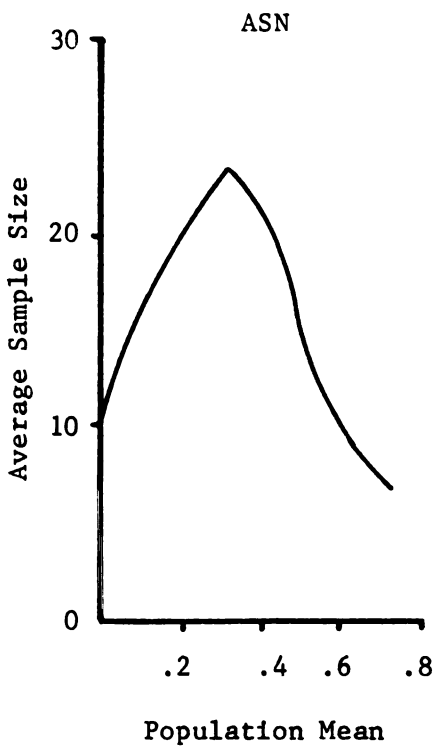


Figure 2a. Decision boundaries of BTTEST for forest tent caterpillar egg mass sampling (see text for details).

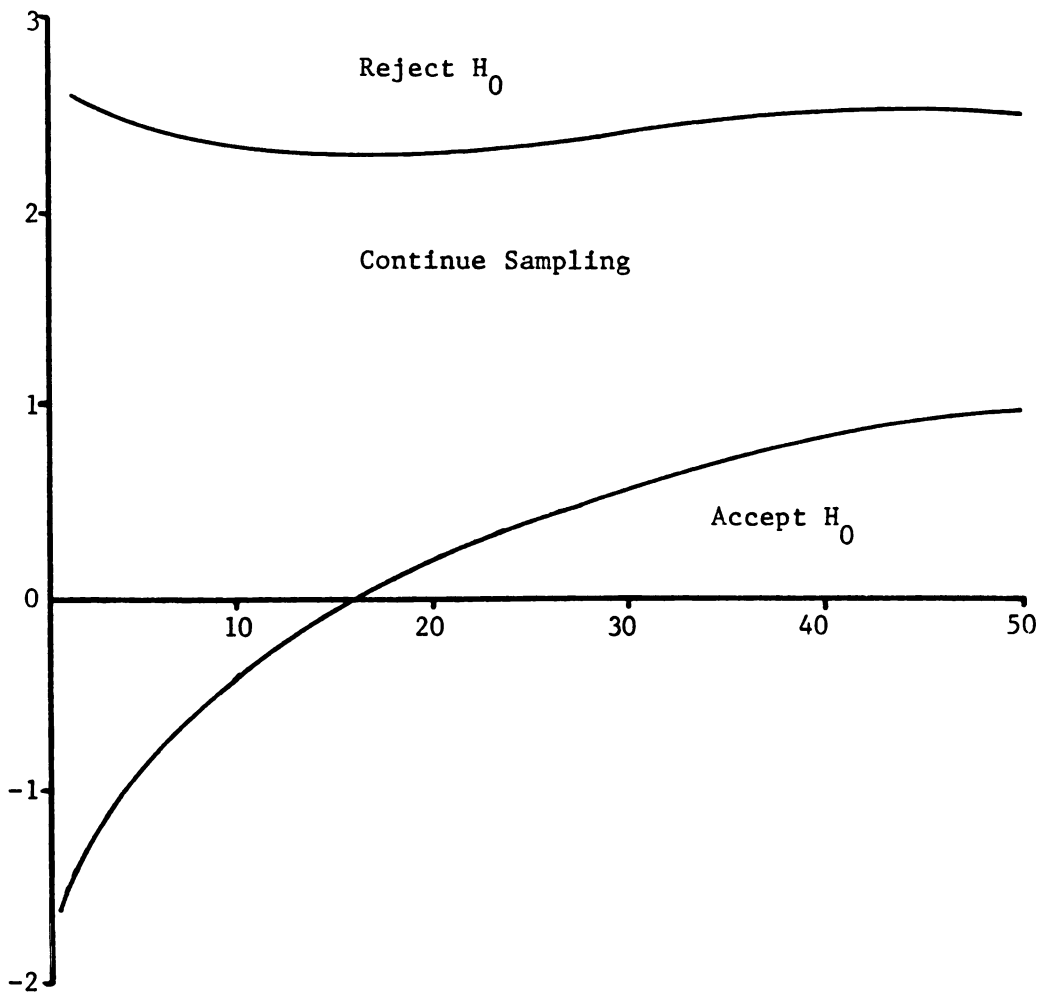


Figure 3a. Decision boundaries of FTTEST for forest tent caterpillar egg mass sampling (see text for details).

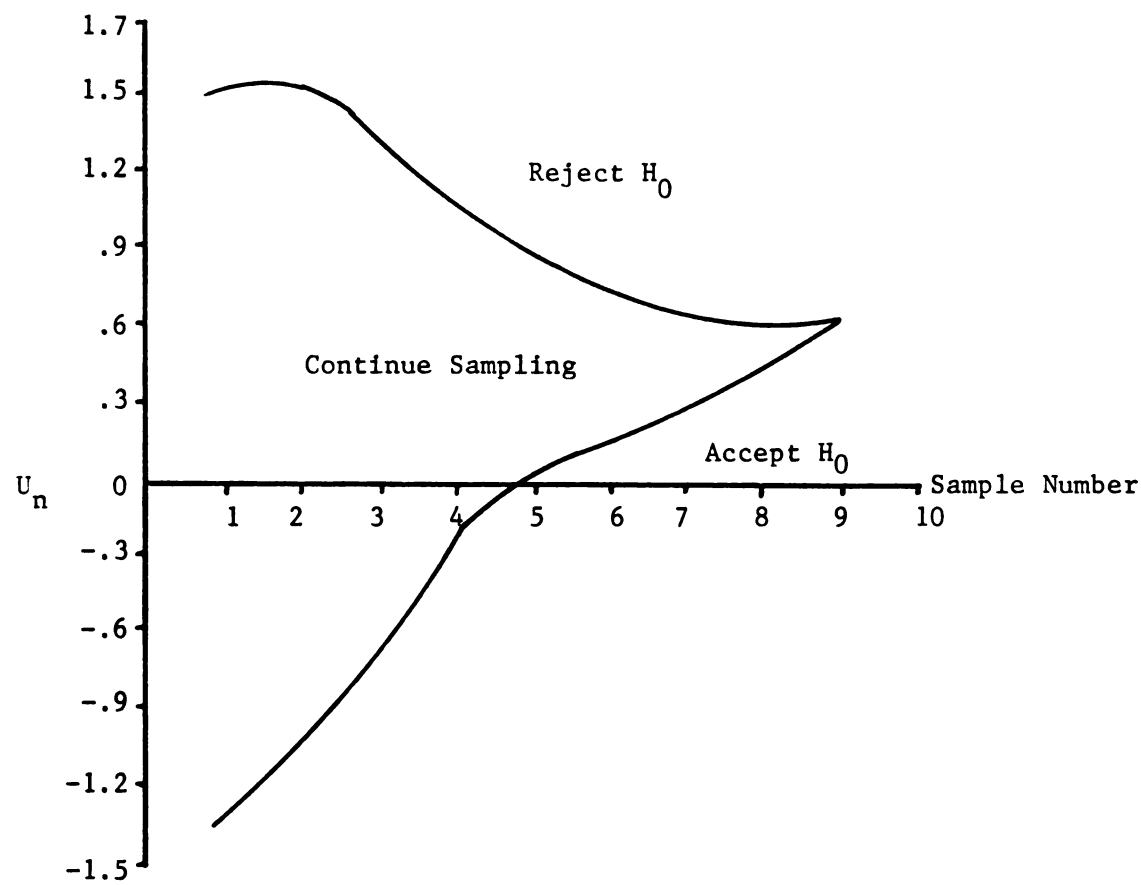
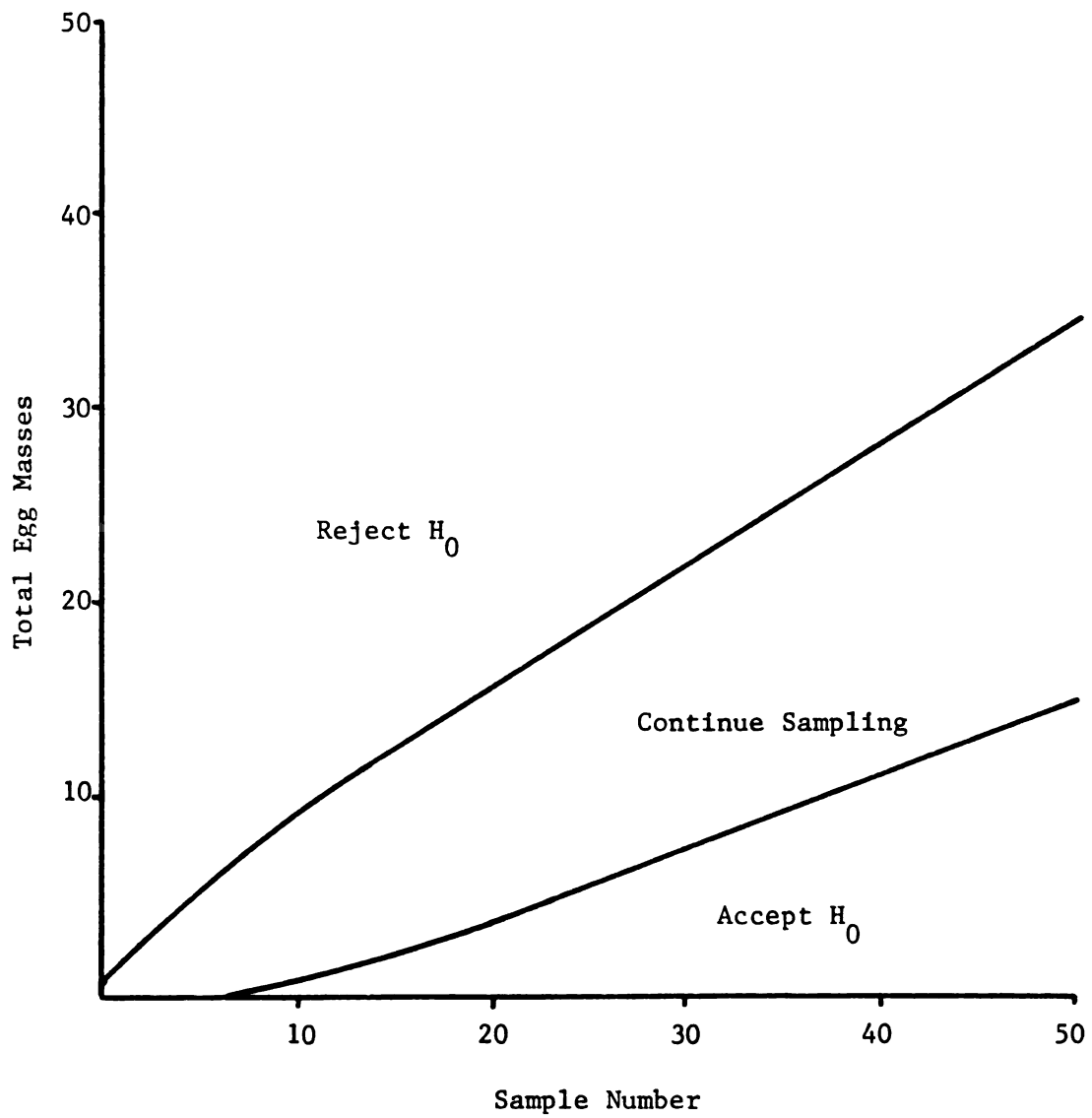


Figure 4a. Decision boundaries of ITEST for forest tent caterpillar egg mass sampling (see text for details).



Subroutine STAT

STAT calculates the mean (MEANEG), variance (VAREGG) and total number of egg masses (TOTEGG), and an estimate of k (KEST) of the negative binomial. Input into the routine consists of the population of 500 branch samples (EGG). A maximum likelihood estimate of k is calculated using an iterative approach until the difference between two successive estimates is 0.001 or twenty five iterations have been processed. Finally, a frequency distribution of egg masses is created.

```
      SUBROUTINE STAT (EGG,MEANEG,KEST,TOTEGG)
      DIMENSION EGG(500), A(6)
      REAL LRB,MEANEG,KEST,MLHL,MLHR
      FO=F1=F2=F3=F4=F5=TOTEGG=TOTEG2=MLHR=0
C CALCULATE THE FREQUENCY DISTRIBUTION
      DO 100 I=1,500
      EGGTT = EGG(I)
      IF (EGGTT.EQ.0.) FO =FO+1
      IF (EGGTT.EQ.1.) F1 = F1+1
      IF (EGGTT.EQ.2.) F2 = F2+1
      IF (EGGTT.EQ.3.) F3 = F3+1
      IF (EGGTT.EQ.4.) F4 = F4+1
      IF (EGGTT.EQ.5.) F5 = F5+1
      TOTEGG = TOTEGG +EGG(I)
      TOTEG2 = TOTEG2 + (EGG(I)**2)
100   CONTINUE
C CALCULATE MEAN AND VARIANCE
      MEANEG = TOTEGG/500
      VAREGG = (TOTEG2 - ((TOTEGG**2)/500))/499
C CALCULATE K ESTIMATE
C MOMENT ESTIMATE OF K
      KEST = (MEANEG**2)/(VAREGG-MEANEG)
C CALCULATE K WITH MAXIMUM LIKELIHOOD METHOD
C A(X) = TOTAL NUMBER OF COUNTS EXCEEDING X
      A(1) = 500-FO
      A(2) = A(1) - F1
      A(3) = A(2) -F2
      A(4) = A(3) -F3
      A(5) = A(4) -F4
      A(6) = A(5) -F5
C
C CALCULATE LEFT PORTION OF MLH EQUATION.
      DO 110 J=1,25
      ADD = .1/J
      MLHL = 500*(ALOG10(1+(MEANEG/KEST)))*2.30259
C CALCULATE RIGHT PORTION OF MLH EQUATION
      MLHR = 0
      DO 120 N=1,6
      MLHR = MLHR + (A(N)/(KEST+(N-1)))
120   CONTINUE
C CALCULATE THIS DIFFERENCE BETWEEN LEFT AND RIGHT PORTIONS
      DIFF = MLHL-MLHR
      IF (ABS(DIFF).LE. .001) GO TO 200
C ADD OR SUBTRACT (ADD) TO K ESTIMATE AND REPEAT
      IF (DIFF.LT.0.) KEST = KEST + ADD
      IF (DIFF.GT.0.) KEST = KEST -ADD
110   CONTINUE
200   CONTINUE
      PRINT 220
220   FORMAT ("OMEAN,  VARIANCE,K,   ITERATIONS, DIFFERENCE")
      PRINT 230, (MEANEG,VAREGG,KEST,J,DIFF)
230   FORMAT (1X,F5.3,4X,F6.4,1X,F8.4,3X,I2,2X,F8.4)
      PRINT 240
```

```
240   FORMAT (" EGGS IN CLASSES 0 THROUGH 5 RESPECTIVELY")
      PRINT 250, (F0,F1,F2,F3,F4,F5)
250   FORMAT (12X,6(F4.0,3X))
      RETURN
      END
```

Subroutine NEGBIN

NEGBIN distributes egg masses according to a negative binomial distribution with exogenously specified mean (MEANEG) and k (KEST). Change in k relative to the mean is given by;

KEST = (estimated k) + (BETA * MEANEG)

where BETA is specified. With each iteration the mean increases linearly through the function;

MEANEG = MEANEG + CHANGE

where CHANGE is specified. The probability of occurrance of egg masses per branch with a range of zero to five egg masses is given P(X) where X is equal to one through six. Egg masses are allocated over a sample of 500.

```
      SUBROUTINE NEGBIN (KEST,EGG,MEANEG,CHANGE, BETA )
      DIMENSION EGG (500), PX(6), SAMP(6)
      REAL KEST,MEANEG
      F0=F1=F2=F3=F4=F5=0
C CALCULATE CHANGE IN THE MEAN AND KEST
      MEANEG = MEANEG + CHANGE
      KEST = 1.0768 + (BETA*MEANEG)
C CALCULATE P(X=0) = PX(1)
      PX(1) = 1/((1+(MEANEG/KEST))**KEST)
C CALCULATE P(X(GT)0) WITH GENERAL FORM:
C P(X=J) = P(X=J-1)*((K+J-1)/J)*(XBAR/(XBAR+K)
      DO 10 N=2,6
      M=N-1
      PX(N) = PX(M)*((KEST + N-2)/(N-1))*(MEANEG/(MEANEG+KEST))
10    CONTINUE
C CALCULATE NUMBER OF SAMPLES WITH X EGGS PER SAMPLE
C AND TOTAL SAMPLES IN DISTRIBUTION
      COUNT = 0
      DO 20 J=1,5
      N=7-J
      SAMP(N) = PX(N) * 500
      I = (SAMP(N) + .5)
      DO 30 K=1,I
      M = K+COUNT
      EGG(M) = 6-J
30    CONTINUE
      COUNT = COUNT + SAMP(N) + .5
20    COUNT
      SAMP(1) + PX(1) * 500
      J = COUNT
      DO 40 I=J,500
      EGG(I) = 0
40    CONTINUE
C
C
      PRINT 50
50    FORMAT ("0SAMPLES DEFINED IN DISTRIBUTION FOR X; 0-5")
      PRINT 60, (SAMP(J), J=1,6)
60    FORMAT (1X,5(3X,F7.3))
      PRINT 70
70    FORMAT (" MEAN OF SAMPLE, KEST OF SAMPLE")
      PRINT 80, MEANEG,KEST
80    FORMAT (2X,F7.4,2X,F7.4)
      DO 100 I=1,500
      EGGTT = EGG(I)
      IF (EGGTT.EQ.0) F0 = F0+1
      IF (EGGTT.EQ.1.) F1 = F1+1
      IF (EGGTT.EQ.2.) F2 = F2+1
      IF (EGGTT.EQ.3.) F3 = F3+1
      IF (EGGTT.EQ.4.) F4 = F4+1
      IF (EGGTT.EQ.5.) F5 = F5+1
100   CONTINUE
      PRINT 200
```

```
200   FORMAT (" SAMPLES IN CLASSES 0 THROUGH 5 RESPECTIVELY")
      PRINT 210, F0,F1,F2,F3,F4,F5
210   FORMAT (12X,6(F4.0,3X))
      RETURN
      END
```

Subroutine SEQUAN

SEQUAN randomly samples the egg mass population 1000 times with a truncation point of 50 observations per sample and classifies the population according to the Sequential Probability Ratio Test (Wald 1947). Input consists of the egg mass population (EGG). Output consists of the number of rejections (NREJ) and acceptances (NACEPT) of the null hypothesis, no decision cases (NODEC) and the average sample size (AVGSAM). The upper and lower rejection boundaries (URB and LRB) are given by;

URB = 3.125 + (.323*N)

LRB = -3.125 + (.323*N)

for α and β equal to 0.10 where N is the sample size. For α and β equal to 0.05 the equations take the form;

URB = 4.193 + (.323*N)

LRB = -4.193 + (.323*N)

```
      SUBROUTINE SEQUAN (EGG)
      DIMENSION EGG (500)
      REAL LRB
      NREJ = NACEPT = TOTSAM = 0
      DO 500 K= 1,1000
      EGGT = 0
C WITH AN OPEN TEST A MAX OF 50 SAMPLES ARE TAKEN
      DO 60 N=1,50
      TOTSAM = TOTSAM +1
C GENERATE A RANDOM SAMPLE
      I= 500*RANF(0.) +1
      EGGT = EGGT + EGG(I)
C CALCULATE URB AND LRB
C URB = IU + B(N)
      URB = 3.125 + (.323*N)
C LRB = IC + B(N)
      LRB = -3.125 + (.323*N)
C DECISION MAKING
      IF (EGGT .GT. URB) GO TO 100
      IF (EGGT .LT. LRB) GO TO 110
60    CONTINUE
      GO TO 500
100   NREJ = NREJ +1
      GO TO 500
110   NACEPT = NACEPT +1
500   CONTINUE
      NODEC = 1000 - NREJ - NACEPT
      AVGSAM = TOTSAM/1000
C PRINT RESULTS
      PRINT 200
200   FORMAT (" SEQUAN SUMMARY:ACCEPTANCES, REJECTIONS, NO DECISIONS
      PRINT 250, (NACEPT,NREJ,NODEC)
250   FORMAT (2X,3(I3,8X))
      PRINT 300
300   FORMAT (" AVERAGE NUMBER OF SAMPLES PER DECISION")
      PRINT 350, AVGSAM
350   FORMAT (3X,F7.4)
      RETURN
      END
```

Subroutine BTTEST

BTTEST randomly samples the egg mass population 1000 times with a truncation point of 50 observations per sample and classifies the population according to Barnard's sequential t-test (Barnard 1952). Input consists of the egg mass population (EGG). Upper and lower rejection boundaries (UNR, UNA) are initialized as these were determined from published tables (National Bureau of Standards t-test Tables).

The test initially requires seven random samples. Following these samples the decision statistic (DESTAT) is calculated and recalculated for each subsequent single sample. The first and subsequent test statistics are calculated as;

$$DESTAT = TDEV/(TDEVSQ)^{\frac{1}{2}}$$

where; 1) TDEV = ΣDEV and DEV is the deviation from the critical density of each observation, 2) TDEVSQ = $\Sigma(DEV)^2$. Output consists of the number of acceptances (NACEPT) and rejections (NREJ) of the null hypothesis, no decision cases (NODEC) and average sample number (AVGSAM).

```
      SUBROUTINE BTTEST (EGG)
      DIMENSION UNR(43), UNA(43), EGG(500)
      NREJ = NACEPT = TSAMP = 0
C ENTER VALUES OF UNR, UNA, A=B = .05, D=.5
C
      DATA UNA /-1.51,-1.33,-1.15,-1.034,-.918,-.802,-.686,-.57,-.498
     +-.426,-.354,-.282,-.21,-.154,-.098,-.042,.014,.07,.114,.158,.202,
     +.246,.29,.328,.366,.404,.442,.48,.514,.548,.582,.616,.65,.678,
     +.706,.734,.762,.79,.816,.842,.868,.894,.92/
      DATA URN /2.56,2.51,2.46,2.436,2.412,2.388,2.364,2.34,2.334,2.328,
     +2.322,2.316,2.31,2.308,2.306,2.304,2.302,2.30,2.304,2.308,2.312,
     +2.316,2.32,2.328,2.336,2.344,2.352,2.36,2.368,2.376,2.384,2.392,
     +2.4,2.408,2.416,2.424,2.432,2.44,2.45,2.46,2.47,2.48,2.49/
      DO 100 M=1,1000
      TDEV = TDEVSQ = 0
C THE TEST REQUIRES 7 RANDOM SAMPLES INITIALLY
      DO 110 J=1,7
      I= 500*RANF(0.) +1
C CALCULATE DEVIATION, SQUARED,TOTALS,FROM HO
      DEV = EGG(I) - .499999
      DEVSQ = DEV**2
      TDEV = TDEV + DEV
      TDEVSQ = TDEVSQ + DEVSQ
110   CONTINUE
      TSAMP = TSAMP + 7
C CALCULATE DECISION STATISTICS FOR N UP TO 50
      DO 120 L = 1,43
      TSAMP = TSAMP + 1
      I = 500*RANF(0.) +1
      DEV = EGG(I) - .499999
      DEVSQ = DEV**2
      TDEV = TDEV + DEV
      TDEVSQ = TDEVSQ + DEVSQ
      IF (TDEVSQ .EQ.0.) GO TO 200
C CALCULATE DECISION STATISTIC
      DESTAT = TDEV/(SQRT(TDEVSQ))
C DECISION MAKING
      IF (DESTAT .GT. UNR(L)) GO TO 300
      IF (DESTAT .LT. UNA(L)) GO TO 310
      GO TO 120
200   DESTAT = 0
      IF (DESTAT .LT. UNA (L)) GO TO 310
120   CONTINUE
      GO TO 100
300   NREJ = NREJ +1
      GO TO 100
310   NACEPT = NACEPT + 1
100   CONTINUE
      AVGSAM = TSAMP/1000
      NODEC = 1000-NREJ-NACEPT
C PRINT RESULTS
      PRINT 400
400   FORMAT ("-SUMMARRY OF BARNARD'S TTEST DECISIONS")
      PRINT 410
```

```
410   FORMAT (" ACCEPTANCES, REJECTIONS, NO DECISIONS")
      PRINT 420, NACEPT, NREJ, NODEC
420   FORMAT (5X,I4,5X,I4,8X,I4)
      PRINT 430
430   FORMAT (" AVERAGE SAMPLE NUMBER")
      PRINT 440, AVGSAM
440   FORMAT (4X,F6.2)
      RETURN
      END
```

Subroutine TTEST

TTEST randomly samples the egg mass population 1000 times and classifies the population according to Fowler's and O'Regan's truncated sequential t-test (Fowler and O'Regan 1974). Input consists of the egg mass population (EGG) and the number of branches (NN) to serve as an observation. This is refered to as sample surface in the text. Upper and lower rejection boundaries (UNR, UNA) are initialized.

Two random samples are initially required. The decision statistic (DESTAT) is calculated following this sampling and for each subsequent single observation until a decision is reached through;

DESTAT = TOTSUM/TOTSSQ

where; 1) TOTSUM = ΣSUM and SUM is the deviation from the critical density for each sample, 2) TOTSSQ = $\Sigma(SUM)^2$. Output consists of the number of acceptances (NACEPT) and rejections (NREJ) of the null hypothesis and average number of branches sampled (AVGBRN).

```
      SUBROUTINE TTEST (EGG, NN)
      DIMENSION UNR(10), UNA(10), EGG(500)
      NREJ = NACEPT = 0
      DATA UNR / 1.4093,1.544,1,265,1,036,.8958,.7795,.7483,.7149,.7385
      DATA UNA / -1.3261,-.9465,-.5178,-.2148,.0199,.2343,.4265,.5772,
     +.7385
      TSAMP = 0
      DO 500 M = 1,1000
C THE TEST REQUIRES 2 RANDOM SAMPLES INITIALLY
      EGG1 = EGG2 = 0
      DO 20 N = 1,NN
      I = 500*RANF(0.)+1
      EGG1 = EGG1 + EGG(I)
      I = 500*RANF(0.)+1
      EGG2 = EGG2 + EGG(I)
20    CONTINUE
      TSAMP = TSAMP + 1
C CALCULATE SUM AND SUM SQUARED EGG
      SUM1 = EGG1 - (.49999*NN)
      SUM2 = EGG2 - (.49999*NN)
      SUMSQ1 = SUM1**2
      SUMSQ2 = SUM2**2
C CALCULATE TOTAL SUM, TOTAL SUMSQUARED
      TOTSUM = SUM1 + SUM2
      TOTSSQ = SUMSQ1 + SUMSQ2
C CALCULATE DECISION STATISTIC
      DESTAT = TOTSUM/(SQRT(TOTSSQ))
      IF (DESTAT.GT.UNR(1)) GO TO 100
      IF (DESTAT.LT.UNA(1)) GO TO 110
C IF NO DECISION WAS MADE CONTINUE SAMPLING, MAX N = 10
      DO 10 J=1,8
      TSAMP = TSAMP+1
      EGGS = 0
      DO 30 N=1,NN
      I = 500*RANF(0.)+1
      EGGS = EGGS + EGG(I)
30    CONTINUE
      L = J+1
      SUM = EGGS - (.49999*NN)
      SUMSQ = SUM**2
      TOTSUM = TOTSUM + SUM
      TOTSSQ = TOTSSQ + SUMSQ
      DESTAT = TOTSUM/(SQRT(TOTSSQ))
      IF (DESTAT.GT.UNR(L)) GO TO 100
      IF (DESTAT.LT.UNA(L)) GO TO 110
10    CONTINUE
      GO TO 500
100   NREJ = NREG + 1
      GO TO 500
110   NACEPT = NACEPT + 1
500   CONTINUE
      TBRAN = TSAMP * NN
```

```
      AVGBRN = TBRAN/1000
C PRINT RESULTS
      PRINT 200
200   FORMAT (" SUMMARY OF T-TEST DECISIONS")
      PRINT 210
210   FORMAT (" REJECTIONS, ACCEPTANCES, SAMPLE SURFACE")
      PRINT 220, NREJ, NACEPT, NN
220   FORMAT (2X,3(I3,8X))
      PRINT 230
230   FORMAT (" AVERAGE NUMBER OF BRANCHES SAMPLED PER DECISION")
      PRINT 240, AVGBRN
240   FORMAT (5X,F7.4)
      RETURN
      END
```

Subroutine SEQRT

SEQRT randomly samples the egg mass population 1000 times with a truncation point of 50 observations per sample and classifies the population according to Iwao's sequential test (Iwao 1975). Input consists of the egg mass population (EGG) and critical density (CDEN) which is to be exceeded for rejection of the null hypothesis. The upper (URB) and lower (LRB) decision boundaries take the form;

URB = (N*CDEN) + A

LRB = (N*CDEN) - A

where; A = (T * (0.9948*N*CDEN)+(0.6952*N*CDEN^2) and T is the desired confidence probability (Student's t). Output consists of the number of rejections (NREJ) and acceptances (NACEPT) of the null hypothesis, no decision cases (NODEC) and average sample size (AVGSAM).

```
      SUBROUTINE SEQRT (EGG, CDEN)
C CALCULATE DECISION BOUNDARIES; N=1, (1), 50
C DECISION BOUNDARY FORM; NM + T(SQRT(N(A+1)M + (B-1)M**2))
C M=CDEN = CRITICAL DENSITY, T = STUDENT'S T
      DIMENSION URB(50), LRB (50), EGG(500)
      REAL LRB
      T = 1.645
      DO 10 N=1,50
      A = T*(SQRT((0.9948*N*CDEN) + (.6952*N*(CDEN)**2)))
      LRB(N) = (N*CDEN) - A
      URB(N) = (N*CDEN) + A
10    CONTINUE
      NGTHO = NLTHO = TOTSAM = 0
C CALCULATE DECISIONS STATISTIC FOR 1000 SAMPLES
      DO 500 K=1,1000
      EGGT = 0
C A MAXIMUM OF 50 SAMPLES ARE TAKEN
      DO 60 N=1,50
      TOTSAM = TOTSAM+1
C GENERATE A RANDOM SAMPLE
      I = 500*RANF(0.)+1
      EGGT = EGGT + EGG(I)
C DECISION MAKING
      IF (EGGT.GT.URB(N)) GO TO 100
      IF (EGGT.LT.LRB(N)) GO TO 110
60    CONTINUE
      GO TO 500
100   NGTHO = NGTHO + 1
      GO TO 500
110   NLTHO = NLTHO + 1
500   CONTINUE
      NODEC = 1000 - NGTHO - NLTHO
      AVGSAM = TOTSAM/1000
      AVGDEC = (TOTSAM -(NODEC*50))/(1000-NODEC)
C PRINT RESULTS
      PRINT 200
200   FORMAT (" SEQRT: GREATER THAN HO, LESS THAN HO, NO DECISION")
      PRINT 250, NGTHO,NLTHO,NODEC
250   FORMAT (3(10X,I4))
      PRINT 300
300   FORMAT (" AVERAGE SAMPLES, AVERAGE SAMPLES W/ DECISION")
      PRINT 350, AVGSAM, AVGDEC
350   FORMAT (4X,F7.3,11X,F7.3)
      RETURN
      END
```

REFERENCES CITED

References Cited

Abrogast, R. T., and M. A. Mullen. 1978. Spatial distribution of eggs by ovipositing Indian meal moths, Plodia interpunctella (Hubner) (Lepidoptera: Pyralidae). Res. Popul. Ecol. 19:148-154.

Baker, A. G. 1950. Properties of some tests in sequental analysis. Biometrika. 37:334-342.

Barnard, G. A. 1952. The frequency justification of certain sequential tests. Biometrika. 39:144-150.

Berthet, P., and G. Gerard. 1965. A statistical study of the micro-distribution of Oribatei (Acari). Part I: The distribution pattern Oikos. 16:214-227.

Bliss, C. I., and A. R. G. Owen. 1958. Negative binomial distribution with a common k. Biometrics. 45:37-58.

Breyer, K. A. 1968. On the distribution of cattle grubs (Diptera: Hypodermatidae) in the herds of cattle. Parasitology. 2:322-333.

Connola, D. P., W. E. Waters, and W. E. Smith. 1957. The development and application of a sequential sampling plan for forest tent caterpillar in New York. New York State Museum and Science Service Bulletin 366. 22p.

Davies, O. L. 1956. Design and Analysis of Industrial Experiments. Hafner Publishers, New York. 635p.

Elliot, J. M. 1977. Some Methods for the Statistical Analysis of Samples of Bethnic Invertebrates. Freshwater Biological Association Scientific Publication No. 25. 160p.

Fowler, G. A., and W. G. O'Regan. 1974. One-sided truncated sequential t-test: Application to natural resource sampling. USDA Forest Service Research Paper, PSW-100. 17p.

Harcourt, D. G. 1961. Spatial pattern of the imported cabbageworm, Pieris rapae (L.) (Lepidoptera: Pieridae), on cultivated Cruciferae. Can. Ent. 93:945-952.

Harcourt, D. G. 1963. Population dynamics of Leptinotarsa decimlineata (Say) in eastern Ontario I. Spatial pattern and transformation of field counts. Can. Ent. 95:813-820.

Harcourt, D. G. 1964. Spatial pattern of the cabbage looper, Trichoplusia ni, on crucifers. Ann. Ent. Soc. Amer. 58:89-94.

Iwao, S. 1968. A new regression method for analyzing the aggregation pattern of animal populations. Res. Popul. Ecol. 10:1-20.

Iwao, S. 1975. A new method of sequential sampling to classify populations relative to a critical density. Res. Popul. Ecol. 16:281-288.

Iwao, S., and E. Kuno. 1968. Use of the regression of mean crowding on mean density for estimating sample size and the transformation of data for the analysis of variance. Res. Popul. Ecol. 10:210-214.

Iwao, S., and E. Kuno. 1971. An approach to the analysis of aggregation pattern in biological populations. Statistical Ecology, Vol. 1. Penn. State Univ.

Kobayashi, S. 1968. Population density and drgree of aggregation with special reference to a common k in the negative binomial distribution. Japanese J. Ecol. 9:113-121.

Kuno, E. 1963. A comparative analysis of the distribution of nymphal populations of some leaf and planthoppers on rice. Res. Popul. Ecol. 5:31-43.

Marshall, P. T., and W. H. Hoffard. 1977. Biological evaluation following two years of forest tent caterpillar defoliation in south central Indiana-1976. USDA Forest Service Evaluation Report, D-8-77. 23p.

Lloyd, M. 1967. Mean Crowding. J. Anim. Ecol. 36:1-30.

Neyman, J., and E. S. Pearson. 1928. On the use and interpretation of certain test criteria for purposes of statistical inference. Biometrika. 20A:175-183.

Onsager, J. A. 1976. The rationale of sequential sampling, with emphasis on its use in pest management. USDA Technical Bulletin No. 1526. 19p.

Pearson, E. S., and N. W. Please. 1975. Relation between the shape of population distribution and the robustness of four simple test statistics. Biometrika. 62:223-241.

Pielou, E. C. 1977. Mathematical Ecology. John Wiley and Sons, New York. 385p.

Pieters, E. P. 1978. Bibliography of sequential sampling plans for insects. Bull. Ent. Soc. Amer. 24:372-374.

Rushton, S. 1950. On a sequential t-test. Biometrika. 37:326-333.

Shiyomi, M., K. Nakamura, and M. Vemura. 1976. A rapid estimation of animal population density on the assumption of negative binomial distribution. Res. Popul. Ecol. 18:28-38.

Stevens, L. M., J. V. McQuire, and A. L. Steinhouer. 1976. Simulation and sampling of the negative binomial distribution with special emphasis on the parameter k as applied to the alfalfa weevil. Univ. of Maryland Ag. Exp. Sta. MP 874.

Wadley, F. M. 1950. Notes on the form of distribution of insect and plant populations. Ann. Ent. Soc. Amer. 43:581-586.

Wald, A. 1947. Sequential Analysis. John Wiley and Sons, Inc. New York. 212p.

Waters, W. E. 1955. Sequential sampling in forest insect surveys. For. Sci. 1:68-79.