M.S. THESIS PROBLEM:

TESTS FOR MORTALITY IN THE k-SAMPLE TAG-RECAPTURE EXPERIMENT

by

BU-381-M D. S. Robson July, 1971

Abstract

In a k-sample tag-recapture experiment where the untagged elements in the i^{th} sample are given tags before the entire sample of n_i elements is returned to the population, the estimation procedure used is dependent upon the assumed nature of the population dynamics in the interims between samples. The hypothesis H_0 that no change occurs in the population during the course of the experiment may be tested against the alternative hypothesis H_1 that mortality occurs between sampling dates by defining a critical region in the space of the statistic S_1 which is sufficient with respect to H_1 such that this critical region has size α with respect to the conditional measure $P_{H_0}(S_1|S_0)$, where S_0 is sufficient with respect to H_0. In this instance S_1 is the sequence

R_i = number of elements in the i^{th} sample that are later recaptured at least once

and

$$S_0 = \sum_1^k (n_i - R_i).$$

The conditional distribution $P_{H_0}(S_1|S_0)$ is not expressible in closed form, and the problem is to find an adequate, large sample approximation to this distribution and then construct an approximate test procedure.

Paper No. BU-381-M in the Biometrics Unit Mimeograph Series, Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York 14850.

M.S. THESIS PROBLEM:

TESTS FOR MORTALITY IN THE k-SAMPLE TAG-RECAPTURE EXPERIMENT

by

BU-381-M D. S. Robson July, 1971

Introduction

In a k-sample tag-recapture experiment on a population closed to recruitment or other forms of immigration the population size N_i at the time the i^{th} sample is drawn is estimated by $R_i C_i / n_i$, where

n_i = size of the i^{th} sample

R_i = number of elements in the i^{th} sample that are later recaptured at least once

C_i = number of distinct elements in all samples after the i^{th}

$$= \sum_{j=i+1}^{k} (n_i - R_i).$$

If the population is also closed to mortality during the sampling period then (see BU-348-M) population size N is estimated implicitly by the equation

$$C_o = \hat{N}\left[1 - \prod_{1}^{k} (1 - \frac{n_i}{\hat{N}})\right] .$$

When the question of mortality is in doubt, a statistical test may be devised by defining a critical region in the space of the sufficient statistic $\underset{\sim}{R} = (R_1, \cdots, R_{k-1})$ which has size α with respect to the conditional distribution $P_{H_o}(\underset{\sim}{R}|C_o)$ under the hypothesis of no mortality. A derivation of $P_{H_o}(\underset{\sim}{R}|C_o)$ is outlined in this preliminary report.

Paper No. BU-381-M in the Biometrics Unit Mimeograph Series, Department of Plant Breeding and Biometry, Cornell University, Ithaca, New York 14850.

Conditional Distribution of the Recapture Vector

Under the hypothesis H_1 that population size declines as a result of mortality, $N_1 \geq \cdots \geq N_k$, and that mortality is random with respect to previous capture history, the joint probability distribution of $\underset{\sim}{R} = (R_1, \cdots, R_{k-1})$ may be expressed as

$$P_{H_1}(\underset{\sim}{R}|\underset{\sim}{N}) = \prod_1^{k-1} P_{H_1}(R_i|C_i, \underset{\sim}{N})$$

$$= \prod_1^{k-1} \binom{n_i}{R_i}\binom{N_i - n_i}{C_i - R_i} \Big/ \binom{N_i}{C_i}$$

and since $H_0 \subset H_1$ then

$$P_{H_0}(\underset{\sim}{R}|N) = \prod_1^{k-1} \binom{n_i}{R_i}\binom{N - n_i}{C_i - R_i} \Big/ \binom{N}{C_i} \,. \tag{1}$$

Since C_0 is a possible value of N and since

$$R_{k-1} = \sum_1^k n_i - C_0 - \sum_1^{k-2} R_i$$

then the joint (non-singular) conditional distribution of $R_1, \cdots, R_{k-2}$ is proportional to the distribution (1) with N replaced by C_0,

$$P_{H_0}(R_1, \cdots, R_{k-2}|C_0) \sim \prod_1^{k-1} \binom{n_i}{R_i}\binom{C_0 - n_i}{C_i - R_i} \Big/ \binom{C_0}{C_i}$$

in fact,

$$P_{H_0}(R_1, \cdots, R_{k-2}|C_0) = \left. \frac{\prod_1^{k-1} \binom{n_i}{R_i}\binom{N - n_i}{C_i - R_i} \Big/ \binom{N}{C_i}}{P_{H_0}(C_0)} \right|_{N = C_0} \,. \tag{2}$$

The denominator of (2) is the convolution of (1) and is not expressible in closed form. For small samples the distribution (2) can be readily calculated in order to tabulate critical values of $(R_1, \cdots, R_{k-1})$ but a large sample approximation to (2) is clearly required to produce results of practical value. Exact results for small samples may then be compared with those obtained from large sample approximations to determine the sample size at which the approximation becomes sufficiently accurate. The test procedure would appear to entail a sequence of one-tailed conditional tests on R_1, R_2, $\cdots$, R_{k-2}, successively.