method was discussed in conjunction with estimation in Model M_b. The context here, however, is different since animals are not reintroduced into the population after initial capture. Zippin (1956) showed that the joint distribution of $\{u_1, u_2, \ldots, u_t\}$, where u_j represents the number of animals removed on the jth occasion, can be written as

$$P[\{u_1, u_2, \ldots, u_t\}] = \frac{N!}{\left[\prod_{j=1}^{t} u_j!\right](N - M_{t+1})!} p^{M_{t+1}} \cdot (1 - p)^{tN - \sum_{j=1}^{t} (t - j + 1)u_j},$$

where p is the probability of removal of any animal on any trapping occasion. Because the parameter space is 2 dimensional (N and p are the only parameters) and the minimal sufficient statistic

$$\left\{M_{t+1}, \sum_{j=1}^{t} ju_j\right\}$$

is 2 dimensional, both parameters are identifiable and may be estimated by maximum likelihood. Such estimators are derived in Appendix D, and are the same as those used to estimate N and p in Model M_b.

An estimator of population size N based on the above model will clearly be unsatisfactory if all animals present do not have an equal probability of removal on a given trapping occasion. In removal experiments, unequal capture probabilities can be caused by time variation or heterogeneity or both. Behavioral variation is nonexistent since members of the population are removed after first capture. We maintain that proper planning and design of the removal experiment can be used to control or reduce time variation (see STUDY DESIGN). Thus, the main problem lies in dealing with heterogeneity of capture (removal) probabilities. In the following development, it is argued that the estimation procedure previously proposed for Model M_{bh} is also appropriate for estimation in removal models. Such a procedure represents a generalization of Moran's (1951) and Zippin's (1956, 1958) constant probability removal model to the case in which heterogeneous probability of removal exists. If the set of removal probabilities p_i, $i = 1, 2, \ldots, N$, is assumed to be a random sample of size N from some probability distribution $G(p; \boldsymbol{\theta})$ parameterized by $\boldsymbol{\theta}$ and defined on [0, 1], then the distribution of the vector of removals $\{u_1, u_2, \ldots, u_t\}$ can be written as

$$P[\{u_1, u_2, \ldots, u_t\}] = \frac{N!}{\left[\prod_{j=1}^{t} u_j!\right](N - M_{t+1})!} \{E[p]\}^{u_1} \{E[(1 - p)p]\}^{u_2} \cdots \{E[(1 - p)^{t-1}p]\}^{u_t} \{E[(1 - p)^{t}\}^{N - M_{t+1}},$$

where

$$E[(1 - p)^{j-1}p] = \int_0^1 (1 - p)^{j-1}p \, dG\,(p; \boldsymbol{\theta}), \quad j = 1, \ldots, t,$$

$$E[(1 - p)^{t}] = 1 - \sum_{j=1}^{t} E[(1 - p)^{j-1}p].$$

This distribution is identical to the distribution of the removals $\{u_1, u_2, \ldots, u_t\}$ under the conditions discussed in the section on Model M_{bh}. Thus, it is easily argued (cf. Appendix J) that the generalized removal method developed for Model M_{bh} is also appropriate for estimation in removal models in which it is assumed that heterogeneity of first capture probability exists. Details of this estimation technique are given in Appendix H, along with an estimator of the asymptotic variance of the population estimator $\hat{N}_R$ (the R stands for "removal," but we note that $\hat{N}_R$ is mathematically the same as $\hat{N}_{bh}$) and a "failure" criterion that must not hold if parameters are to be validly estimated. This criterion ensures that a sufficient decline in the population is being effected by the successive removals.

As explained in the material on estimation of N in the section on Model M_{bh}, our procedure is to look at a sequence of removal models. These correspond to a sequence of assumptions as follows: all $\bar{p}_j$ are equal, or $\bar{p}_1 \neq \bar{p}_2$ but $\bar{p}_2 = \bar{p}_3 = \cdots = \bar{p}_t$, or $\bar{p}_1 \neq \bar{p}_2 \neq \bar{p}_3$, but $\bar{p}_3 = \bar{p}_4 = \cdots = \bar{p}_t$, and so forth. The most general model allows all $\bar{p}_j$ to be different, but this most general model does not allow estimation of N. We define the specific removal model M_{Rk} as the model in which the last $t - k + 1$ values of $\bar{p}_j$ are the same, and for $k > 1$, the first $\bar{p}_1$ to $\bar{p}_k$ are different.

The estimation procedure is to test the goodness of fit of the removal models sequentially from M_{R1} to $M_{R,t-2}$. Thus, we are first testing whether the constant probability model fits. If it does, we use the corresponding maximum likelihood estimator of N. If this simple model is rejected by the chi-square test (at the 20% significance level in program CAPTURE), we then examine the case (model M_{R2}) of $\bar{p}_1 \neq \bar{p}_2$, but $\bar{p}_2 = \bar{p}_3 = \cdots = \bar{p}_t$. The estimator of N, $\hat{N}_R$, used in this generalized removal approach is taken as the ML estimator of N for the selected removal model.

Simulation Results

Bias and Confidence Intervals

A discussion of the bias of the estimator of N associated with the generalized removal method and of the usefulness of its associated confidence intervals was presented in the material on Model M_{bh}. Briefly, the simulation study revealed that the bias of the estimator of N was in most cases not serious, but that confidence intervals achieved only 50 percent coverage on the average. Although the results and discussion of that section were presented in the context of a capture–recapture experiment, all the material is directly applicable here. Because the 2 estimation procedures are identical, the relevant data in both cases involve only the removal statistics from populations with heterogeneous probabilities of first capture. Denoting the estimator of N as $\hat{N}_R$ is intended to remind the reader that this section deals with strict removal data.

Robustness

In the context of removal experiments, the generalized removal method represents a significant step forward with respect to robust estimation of population size. It should be emphasized, however, that the procedure is designed to be robust to failure of the assumption that all animals have the same probability of first capture, and not to failure of the assumption that sampling effort is uniform over trapping occasions. There is some indication in the simulation results that the method performs adequately when there is no heterogeneity of capture probability but there is nonuniformity in sampling rates over time. The method performs very poorly, however, when both heterogeneity and nonuniform sampling rates are present. First, the percentage of experiments which "fail," as determined by the failure criterion, can be very high. For instance, in the population defined by Trial 1 of Model M_{th}, the experiment failed in every one of 100 simulations. Secondly, when the experiment does succeed, bias is usually significant. Some simulated examples are given in Table 11 which help to substantiate these remarks. Complete simulation results are given in Table N.6.b of Appendix N.

Example

Andrzejewski and Jezierski (1966) reported the results of a study designed to estimate population density of European hare *Lepus europaeus* on experimental hunting grounds in Poland. Hares were captured and removed by driving them into nets surrounding the area. Results of the application of the generalized removal method to the data resulting from the 4 drives (removal occasions) are given in Fig. 9. When $k = 1$ (i.e., all animals are assumed to have the same probability of removal), a poor fit to the data results (chi-square goodness of fit value is 13.5, 2 df, and $p = 0.0014$). For $k = 2$, how-

```
OCCASION          J=     1    2    3    4
TOTAL CAUGHT    M(J)=    0  722  913  982 1018
NEWLY CAUGHT    U(J)=  722  191   69   36


  K       N-HAT      SE(N)          CHI-SQ.    PROB.     ESTIMATED P-BAR(J),J=1,...., 4

  1     1028.21  3.789703          13.150     .0014      .6806   .6806   .6806   .6806
  2     1039.10  7.658751           1.528     .2164      .6948   .5916   .5916   .5916


POPULATION ESTIMATE IS        1039 WITH STANDARD ERROR      7.6588

APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL        1023 TO       1055


HISTOGRAM OF U(J)


FREQUENCY     722  191   69   36
-----------------------------------
EACH * EQUALS   73 POINTS

      730      *
      657      *
      584      *
      511      *
      438      *
      365      *
      292      *
      219      *    *
      146      *    *
       73      *    *    *
-----------------------------------
```

FIG. 9. Example of population estimation under the variable probability removal model with data on European hare from Andrzejewski and Jezierski (1966).

ever, an acceptable fit results (p = 0.2164) and, therefore, the point and interval estimates corresponding to Model M_{R2} are chosen. In particular, the point estimate $\hat{N}_R = 1,039 \pm 7.66$ seems much preferable to the estimate of 1,010 that results from the (commonly used) regression method used by Andrzejewski and Je-

TABLE 11.—COMPUTER SIMULATION RESULTS ILLUSTRATING THE ROBUSTNESS OF THE GENERALIZED REMOVAL ESTIMATOR $\hat{N}_R$ USING DATA GENERATED UNDER OTHER MODELS (ALSO SEE APPENDIX N, TABLES N.2.b, AND N.5.b)

Percent relative bias RB	Ave[$\hat{N}_R$]	Population size N	Number of reps R	Number of occasions t	Data model
− 3.2	387.1	400	100	10	M_t, Trial 6[1]
− 9.0	364.1	400	100	7	M_t, Trial 7
2.6	410.3	400	100	7	M_t, Trial 9
− 0.2	199.7	200	100	5	M_t, Trial 10
− 1.9	272.6	400	100	5	M_{th}, Trial 2
22.3	489.1	400	100	5	M_{th}, Trial 4
−10.3	179.4	200	100	7	M_{th}, Trial 5

[1] For example, data generated under Model M_t were used to estimate N, using the generalized removal method $\hat{N}_R$.

```
OCCASION            J=     1    2    3    4    5
TOTAL CAUGHT     M(J)=     0  181  192  196  201  204
NEWLY CAUGHT     U(J)=   181   11    4    5    3


   K      N-HAT      SE(N)         CHI-SQ.    PROB.      ESTIMATED P-BAR(J),J=1,...., 5

   1     204.00   .2094765         70.820    0.0000      .8160   .8160   .8160   .8160   .8160
   2     206.77   3.697937          1.555     .4596      .8754   .4029   .4029   .4029   .4029
   3     208.43   8.365668          1.021     .3124      .8684   .4010   .3306   .3306   .3306


POPULATION ESTIMATE IS          207 WITH STANDARD ERROR       3.6979

APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL        199 TO          215


HISTOGRAM OF U(J)


FREQUENCY    181   11    4    5    3
--------------------------------------
EACH * EQUALS   19 POINTS

     190      *
     171      *
     152      *
     133      *
     114      *
      95      *
      76      *
      57      *
      38      *
      19      *    *
--------------------------------------
```

FIG. 10. Example of population estimation under the variable probability removal model with aquatic insect data from R. F. Raleigh (pers. comm.).

zierski, because the number of hares actually removed was 1,018. The estimated capture probabilities for Model M_{R2} are $\bar{p}_1 = 0.6948$, and $\bar{p}_2 = \bar{p}_3 = \bar{p}_4 = 0.5916$, a result that supports the idea that heterogeneity is operating.

Example

R. F. Raleigh (pers. comm.) provided some results of a removal experiment involving a species of mayfly *Ephemeroptera*. Mayflies were sampled from 10 randomly placed 0.25-m^2 areas in a section of the Poudre River streambed near Fort Collins, Colorado, with 5 removal occasions at each site. A special benthic aquatic sampler was used in the study. Because these are true removal data, the generalized removal estimation procedure was used to produce point and interval estimates for N (Fig. 10). As in the previous example, there is a poor fit for the simple model with constant capture probability (k = 1). When k = 2, however, a good fit to the data results and an estimate $\hat{N}_R = 207$, corresponding to Model M_{R2} is produced. Note the difference between $\bar{p}_1 = 0.8754$ and $\bar{p} = 0.4029$, indicating that a significant heterogeneity in removal probabilities may exist. The computed confidence interval is somewhat unsatisfactory because its lower limit is less than the number of animals actually seen, and reminds one that in removal and livetrapping experiments, conditions necessary for construction of normal theory confidence intervals are

```
OCCASION           J=     1     2     3     4     5     6     7
TOTAL CAUGHT    M(J)=     0    25    51    66    79    91   104   109
NEWLY CAUGHT    U(J)=    25    26    15    13    12    13     5


   K       N-HAT      SE(N)           CHI-SQ.     PROB.      ESTIMATED P-BAR(J),J=1,...., 7

   1      138.07   14.67943            3.879      .5670       .1980   .1980   .1980   .1980   .1980   .1980   .1980
   2      132.46   13.64926            3.623      .4594       .1887   .2218   .2218   .2218   .2218   .2218   .2218
   3      141.63   25.85191            2.754      .4311       .1765   .2229   .1829   .1829   .1829   .1829   .1829
   4      130.96   20.14541            2.688      .2608       .1909   .2454   .1876   .2339   .2339   .2339   .2339
   5      118.54   10.21073            2.133      .1442       .2109   .2779   .2221   .2474   .3675   .3675   .3675


POPULATION ESTIMATE IS          138 WITH STANDARD ERROR     14.6794

APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL          109 TO          167


HISTOGRAM OF U(J)

FREQUENCY     25   26   15   13   12   13    5
------------------------------------------------
EACH * EQUALS    3 POINTS

       27           *
       24      *    *
       21      *    *
       18      *    *
       15      *    *    *
       12      *    *    *    *    *    *
        9      *    *    *    *    *    *
        6      *    *    *    *    *    *    *
        3      *    *    *    *    *    *    *
------------------------------------------------
```

FIG. 11. Example of population estimation under the variable probability removal model with whitefish data from Ricker (1958:150).

not satisfied. In this case, one should truncate the lower limit to the actual number removed, rather than using a lower limit less than M_{t+1}.

Example

Ricker (1958:150) removed whitefish *Coregonus clupeaformis* from Shakespeare Island Lake in Ontario, Canada, on 7 successive occasions by means of gillnetting. Members of whitefish removed in the 13- to 14-inch (3.3–3.6 cm) length class are shown in Fig. 11, with the results of the generalized removal method. Notice that a good fit to the removals is achieved for $k = 1$, so that Zippin's (1956, 1958) constant probability removal model seems appropriate for estimating N. The population estimate of $\hat{N}_R = 138$ is in close agreement with Ricker's estimate of 136 which was obtained by using DeLury's (1947) regression method assuming equal effort. (We wish to reemphasize that on theoretical grounds the ML estimation procedure is superior to the various regression techniques often used with removal studies.) The 95 percent confidence interval estimate of [109, 167] is very informative. Moreover, the true confidence level of the interval probably is close to the stated level of 0.95, since the estimated probability of removal is nearly 0.20.

Discussion

Conducting a removal experiment for purposes of estimating population size may sometimes prove more feasible than a capture–recapture approach (refer to STUDY DESIGN for further discussion). In such cases, the experimenter has available 2 classes of estimation procedures, the catch–effort techniques usually associated with Leslie and Davis (1939) and

DeLury (1947) or the "removal" techniques first introduced by Moran (1951), refined by Zippin (1956, 1958), and generalized here. It is felt that catch–effort techniques are often not appropriate, either because of the assumptions involved or because the concept of effort may be meaningless in many experimental situations. In those frequent situations, we believe the generalized removal method provides the best approach to estimating population size. The operating characteristics of this method are by no means completely satisfactory, in view of the results concerning confidence interval coverage and the failure of the experiment in some situations. However, the fact that the method fails in a given experiment does at least prevent the use of wildly inaccurate estimates in practice and helps to inform the experimenter that the assumptions of the method are not met and the quality of the experiment needs to be upgraded or the design altered, or both. Nevertheless, the method is the most general now available in the literature that is capable of providing useful results and, with further research, improvements in the method should be forthcoming.

TESTS OF MODEL ASSUMPTIONS

In preceding sections, we have recognized 3 distinct sources of variation in capture probabilities, and have given 8 models corresponding to presence or absence of specific types of variability. No estimator of population size has been derived from 3 of those models (M_{th}, M_{tb}, M_{tbh}). Consequently, we have introduced only 5 estimators that are identified below, along with the models where they should be used. The main purpose of this section is to give a strategy for selecting one of the estimation procedures based on tests of the various underlying assumptions.

Philosophy of the Approach

Our philosophy is that one should first present the most general model structure (i.e., assumptions) possible for the given data. Then a series of specific alternative models should be developed based on specific assumptions that cover the various cases of realistic or theoretical interest. Those alternative models form the basis for both testing of assumptions and estimating parameters. Under FUNDAMENTAL CONCEPTS, we gave the general model framework considered here: population closure is assumed and captures are assumed to be independent events with capture probabilities

$$p_{ij}, j = 1, \ldots, t, i = 1, \ldots, N.$$

Such a model structure is useful only as a conceptual starting point because none of the parameters are estimable. Next, we recognized 3 different sources of variation acting on these capture probabilities: (1) variation over time, (2) behavioral variation as a result of first capture (trap response), and (3) variation over individuals (heterogeneity). We also recognized the various possible combinations of these sources of variation.

An almost infinite variety of very specific models could be constructed. We examined 8 models likely to be generally useful for estimation or testing purposes. Those models can all be expressed in terms of the assumptions about variation in the capture probabilities. If each of the 8 models introduced had its own unique estimation procedure, then selection of an estimator would be synonymous with selection of a model. Such is not the case with only 5 estimators and 8 models. Our strategy is to present a testing sequence designed to lead to selection of the most appropriate model. We define "appropriate model" as the simplest model that "fits" the data. When that selection gives a model such as M_{th} for which there is no theoretically appropriate estimator, we recommend (in some instances) using the next best fitting model for which an estimator exists. It will occasionally be necessary to conclude that no estimate of N can legitimately be calculated from the data.

It is important to recognize that we do

not claim this model selection procedure results in the "true" model. By true model for a capture study we mean that the stated model is an exact representation of the true capture probability structure for that study. Any model is a simplification that can realistically represent only the more predominant features of the study. Also, we must recognize that the most appropriate model will depend upon the amount of data we have. With good data (large grids, many occasions) one might be able to show, for example, that Model M_{bh} was necessary. Yet, if the same study used a much smaller grid, and fewer days, it might be impossible to demonstrate heterogeneity from the data themselves and Model M_b might be indicated as appropriate.

The conceptual goal of our model selection procedure is to achieve an acceptable trade-off between precision and bias. If one uses too simple a model, the estimated population size is likely to be severely biased yet have a deceptively small sampling variance. If the model is complex, but still the wrong one, then estimators are again of uncertain bias. If a model is chosen that is too general, the risk of bias is much reduced, or negligible, but the sampling variance is larger than it needs to be (i.e., we lose precision). Of those 2 cases, the safer one in our opinion is to choose the more general model in hopes of minimizing biases. We state this as our conceptual goal because, not knowing the true model, we can never be certain when analyzing real data that we have made the best choice of a model.

This model selection procedure is based on tests of assumptions of 2 types: (1) tests between specific models, and (2) general goodness of fit tests to specific models. Before discussing these tests we summarize the primary models and associated estimators.

Summary of Models and Estimators

Capture–Recapture Models

Because these models have been described in detail, only a summary is given below, in terms of assumptions and corresponding capture probabilities.

MODEL M_o: $p_{ij} \equiv p$.

Comment: This is the most restrictive model because capture probabilities are assumed to be constant.

Estimation: We recommend the ML estimator (see Appendix B); it does not exist in closed form except for $t = 2$.

MODEL M_t: $p_{ij} \equiv p_j, j = 1, \ldots, t$.

Comment: Capture probabilities vary only with time. This is the model most frequently assumed (often implicitly) in the literature.

Estimation: We recommend the ML estimator (see Appendix C), it does not exist in closed form, except for $t = 2$ (in which case it is the Lincoln estimator). The Schnabel estimator, or variations thereof, is often used with this model, but is only an approximation to the ML estimator.

MODEL M_b: $p_{ij} \equiv \begin{cases} p \text{ for first capture,} \\ c \text{ for recapture.} \end{cases}$

Comment: This is the simplest model of behavioral (trap) response and it has been recognized in the literature (Tanaka 1956).

Estimation: The ML estimator, which is recommended (Appendix D), is a special case of the generalized removal estimator. Zippin's removal model is equivalent to Model M_b for purposes of estimating N, but his suggested estimator is only a close approximation to the true ML estimator.

MODEL M_h: $p_{ij} \equiv p_i, i = 1, \ldots, N$.

Comment: This model allows heterogeneity of capture probabilities over animals, but allows no variation in capture probabilities over time.

Estimation: The jackknife estimator (Appendix E) is recommended when M_h is the underlying model; that estimator was derived specifically for this model. Note, however, that there is no way to derive an optimal estimator for Model M_h if nothing is known about how the capture probabilities vary.

$$\text{MODEL } M_{tb}\text{: } p_{ij} \equiv \begin{cases} p_j \text{ for any first capture, } & j = 1,\ldots,t, \\ c_j \text{ for any recapture, } & j = 2,\ldots,t. \end{cases}$$

Comment: This model allows variation in capture probabilities due to both time and behavior (trap response).

Estimation: Population size N is not estimable under this model. Therefore, there is no satisfactory estimation procedure associated with this model.

MODEL M_{th}: $p_{ij} \equiv p_i p_j$, $j = 1,\ldots,t$, $i = 1,\ldots,N$.

Comment: This is useful as a conceptual model of how time and heterogeneity might operate as a simple product. No estimator can be derived from this model.

Estimation: Population size N is not estimable under this model if straightforward methods are used. Therefore, there is no satisfactory estimation procedure associated with this model

$$\text{MODEL } M_{bh}\text{: } p_{ij} \equiv \begin{cases} p_i \text{ for first capture,} \\ c_i \text{ for all recaptures,} \\ \qquad i = 1,\ldots,N. \end{cases}$$

Comment: This model assumes heterogeneity with trap response, but no time variation as such in the capture probabilities.

Estimation: The generalized removal estimator should perform well here (Appendix H). However, if none of the generalized removal models fit, no estimate can be produced using this model.

$$\text{MODEL } M_{tbh}\text{: } p_{ij} \equiv \begin{cases} p_{ij} \text{ for first capture,} \\ c_{ij} \text{ for all recaptures,} \\ \qquad j = 1,\ldots,t, \\ \qquad i = 1,\ldots,N. \end{cases}$$

Comment: This is a conceptually useful model because it represents the case in which all 3 sources of variation operate.

Estimation: Population size N is not estimable under this model using straightforward methods. Therefore, there is no satisfactory estimation procedure associated with this model.

TABLE 12.—MODELS AND SUGGESTED ESTIMATED PROCEDURES

Model	Theoretically appropriate estimator
M_0	Null ($\hat{N}_0$)
M_t	Darroch ($\hat{N}_t$)
M_b	Zippin ($\hat{N}_b$)
M_h	Jackknife ($\hat{N}_h$)
M_{tb}	(none)
M_{th}	(none)
M_{bh}	Generalized removal ($\hat{N}_{bh}$)
M_{tbh}	(none)

Relationships Between Models

Table 12 summarizes some of the above information. It is important to understand the relationship of the models to one another in order to understand testing between models. One can only truly test one model against another when one model is a special case of the other. For example, Model M_0 is a special case of each of the 3 models M_t, M_b, and M_h. But those 3 models are not related to each other in any simple manner that allows one to choose or test between them with any simple statistical tests.

The relationships among these models is shown in Fig. 12 using an arrow between 2 models to indicate that one is a special case of the other. For example, $M_0 \leftarrow M_t$ means Model M_0 is a special case of Model M_t, because if Model M_t is restricted by the assumption $p_1 = p_2 = \cdots = p_t = p$, Model M_0 results.

The selection of an appropriate capture–recapture model is neither simple nor straightforward and there is, in fact, no optimal or rigorous statistical theory to guide this selection. The jumble of arrows in Fig. 12 is indicative of the difficulty.

Specific Tests to Perform

Before we can give our strategy for model, hence estimator, selection, it is necessary to discuss exactly what hypothesis can be tested concerning variations in capture probabilities. Two types of tests are made:

(1) Specific tests of one model versus a more general model. For example, testing Model M_0 vs. Model M_t tests for time variation in average daily capture probabilities;

(2) General goodness of fit tests of a given model to the data. For example, testing whether Model M_t fits the data amounts to testing whether time variation is the only source of variation in capture probabilities.

There is a basic difference in the nature of the hypotheses associated with those 2 different classes of tests. The objective of the first class of tests is to establish whether or not the more general model produces a significantly better fit to the data than the more restrictive model. On the other hand, the goodness of fit tests of the second class test whether or not the data might reasonably have arisen from the specified model.

Table 13 gives the 7 specific tests we have used in the model selection procedure. The reader should study Table 13 before proceeding. Program CAPTURE has been used to produce an example of the tests. Results are presented in Fig. 13, that is based on simulated data under Model M_0 with parameters $N = 100$, $p = 0.50$, and $t = 5$. The first 3 tests of the model selection procedure separately test each of the simple models M_h, M_b, M_t against Model M_0. These are all chi-square tests (see Appendix K for details). The significance level of the tests is given by the program. For example, in Fig. 13 under the test of Model M_0 vs. Model M_h (test 1) we have a chi-square test statistic of 5.007 (3 degrees of freedom) and the significance level (labeled PROBABILITY OF LARGER VALUE) is 0.17127. Thus, that test is not significant, and we have no basis to believe there is any heterogeneity of capture probabilities.

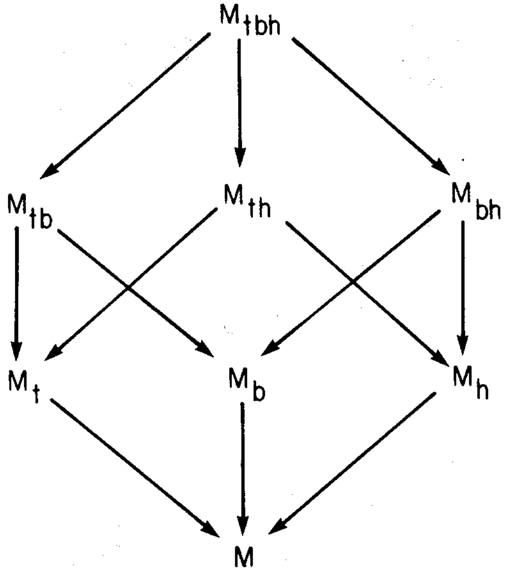


FIG. 12. Relationships among models: arrows indicate that one model is an immediate special case of another model.

Similarly, the simple tests for behavioral variation (Model M_0 vs. Model M_b, test 2) and for time variation (Model M_0 vs. Model M_t, test 3) are not significant in Fig. 13. Those tests had significance levels of 0.98056 and 0.31935, respectively. Based on only these results, we would (correctly, in this case) conclude the appropriate model is Model M_0. When a model more complex than M_0 is the appropriate one, additional tests are needed.

In the second part of the model selection procedure, 4 more tests are presented. The first 3 are simple goodness of fit tests to models M_h, M_b, and M_t. The seventh is a test for behavioral response in the presence of heterogeneity; hence this test helps make a choice between M_b and M_{bh}. That last test is included here because we have an estimator for Model M_{bh}.

The goodness of fit test of Model M_h (test 4) in Fig. 13 results in a chi-square value of 4.556 (4 df) and has a significance

```
OCCASION          J=    1    2    3    4    5
ANIMALS CAUGHT  N(J)=  55   47   54   60   48
TOTAL CAUGHT    M(J)=   0   55   75   88   96   98
NEWLY CAUGHT    U(J)=  55   20   13    8    2
FREQUENCIES     F(J)=  10   38   26   20    4


1. TEST FOR HETEROGENEITY OF TRAPPING PROBABILITIES IN POPULATION.
   NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(H)

          CHI-SQUARE VALUE =     5.007   DEGREES OF FREEDOM =   3   PROBABILITY OF LARGER VALUE =   .17127


2. TEST FOR BEHAVIORAL RESPONSE AFTER INITIAL CAPTURE.
   NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(B)

          CHI-SQUARE VALUE =      .001   DEGREES OF FREEDOM =   1   PROBABILITY OF LARGER VALUE =   .98056


3. TEST FOR TIME SPECIFIC VARIATION IN TRAPPING PROBABILITIES.
   NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(T)

          CHI-SQUARE VALUE =     4.701   DEGREES OF FREEDOM =   4   PROBABILITY OF LARGER VALUE =   .31935

4. GOODNESS OF FIT TEST OF MODEL  M(H)
   NULL HYPOTHESIS OF MODEL  M(H) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(H)

          CHI-SQUARE VALUE =     4.556   DEGREES OF FREEDOM =   4   PROBABILITY OF LARGER VALUE =   .33601

          TEST OF MODEL M(H) BY FREQUENCY OF CAPTURE
          (FREQUENCIES LESS THAN 2T ARE NOT CALCULATED.)

          NUMBER OF CAPTURES  CHI-SQUARE  D.F.  PROBABILITY
          ---------------------------------------------

                  1              5.000     4       .28730
                  2              1.298     4       .86167
                  3              1.949     4       .74519
                  4              5.500     4       .23973

5. GOODNESS OF FIT TEST OF MODEL  M(B)
   NULL HYPOTHESIS OF MODEL  M(B) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(B)

          CHI-SQUARE VALUE =     5.183   DEGREES OF FREEDOM =   6   PROBABILITY OF LARGER VALUE =   .52053

      5A. CONTRIBUTION OF TEST OF HOMOGENEITY OF FIRST CAPTURE PROBABILITY ACROSS TIME

          CHI-SQUARE VALUE =     2.459   DEGREES OF FREEDOM =   3   PROBABILITY OF LARGER VALUE =   .48276

      5B. CONTRIBUTION OF TEST OF HOMOGENEITY OF RECAPTURE PROBABILITIES ACROSS TIME

          CHI-SQUARE VALUE =     2.724   DEGREES OF FREEDOM =   3   PROBABILITY OF LARGER VALUE =   .43610

6. GOODNESS OF FIT TEST OF MODEL  M(T)
   NULL HYPOTHESIS OF MODEL  M(T) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(T)

          CHI-SQUARE VALUE =    59.465   DEGREES OF FREEDOM =  54   PROBABILITY OF LARGER VALUE =   .28338


7. TEST FOR BEHAVIORAL RESPONSE IN PRESENCE OF HETEROGENEITY.
   NULL HYPOTHESIS OF MODEL  M(H)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(BH)

          CHI-SQUARE VALUE =     7.471   DEGREES OF FREEDOM =  10   PROBABILITY OF LARGER VALUE =   .68036

MODEL SELECTION CRITERIA.  MODEL SELECTED HAS MAXIMUM VALUE.

MODEL      M(O)      M(H)      M(B)      M(BH)     M(T)      M(TH)     M(TB)     M(TBH)
CRITERIA   1.00       .96       .26       .51      0.00       .44       .33       .64
```

TABLE 13.—SPECIFIC TESTS OF ASSUMPTIONS USED IN THE MODEL SELECTION PROCEDURE

Test number	Source of variation tested for	Null hypothesis	Alternative hypothesis	Comments
1.	Heterogeneity	Model M_0 fits the data	Model M_h fits the data	This test examines the capture frequencies to see if there is evidence of variability among individual capture probabilities
2.	Trap response after first capture	Model M_0 fits the data	Model M_b fits the data	This is a test for gross behavior effects on capture probabilities
3.	Time variation in capture probabilities	Model M_0 fits the data	Model M_t fits the data	This tests for variation in average daily capture probabilities. Unfortunately the actual test is also sensitive to Model M_b
4.	Trap response and/or time variation given heterogeneity	Model M_h fits the data	Model M_h fails to fit the data	If M_h is the true model we expect this test not to reject. We would also expect test 1 to reject Model M_0 in favor of M_h
4a.	As above	As above	As above	These tests are a specific form of test 4, where for each significantly large capture frequency an individual test can be made for trap response or time variation or both
5.	Heterogeneity and/or time variation given trap response	Model M_b fits the data	Model M_b fails to fit the data	If M_b is the true model we expect this test not to reject. We also would expect test 2 to reject Model M_0 in favor of M_b. This goodness of fit test can be partitioned into two specific tests, useful in their own right (5a, 5b)
5a.	Heterogeneity and/or time variation using first capture only	First capture probabilities are constant	First capture probabilities vary by time and/or animals	This test is identical to the goodness of fit test for the simple removel model
5b.	Heterogeneity and/or time variation using only recaptures	Recapture probabilities are all constant	Recapture probabilities vary by time and/or animals	If M_b is true then this null hypothesis should not be rejected
6.	Trap response and/or heterogeneity given time variation	Model M_t fits the data	Model M_t fails to fit the data	If M_t is the model, then we expect this test to fail to reject, and we expect Model M_0 to be rejected in favor of M_t
7.	Trap response given heterogeneity	Model M_h fits the data	Model M_{bh} fits the data	This test is useful because if we reject M_h in favor of M_{bh} the estimator to use is the generalized removal method

←

FIG. 13. Example of the model selection procedure tests with simulated data. True model is M_0 with parameters $N = 100$, $t = 5$, and $p = 0.5$. Appropriate model is M_0. Suggested estimator is Null.

level of 0.33601. Thus, in this example Model M_h fits the data. This is expected because Model M_0 is a special case of Model M_h, and hence the tests should not reject.

When individual capture frequencies are large enough (we have used the criteria $f_i > 2t$) we have computed a goodness of fit test of Model M_h based on the data for animals captured exactly f_i times. Theoretically, this can be done for all frequencies $f_1, \ldots, f_{t-1}$ that are large enough.

The goodness of fit test of Model M_b (test 5) can also be partitioned into 2 components. Test 5a is actually the goodness of fit test to Zippin's removal model, i.e., it tests whether the probability of first capture is constant over time. In Fig. 13, the test statistic value is 2.459 with significance level 0.48276. Thus, in this case there is no evidence of variation in first capture probability over time. The second component of the test of Model M_b (test 5b) examines for time variation in recapture probabilities. In Fig. 13, the test statistic value is 2.724. The goodness of fit test statistic for Model M_b is the sum of those 2 chi-square values. In Fig. 13, the value is 5.183 (6 df) with significance level 0.52053. Thus, we conclude Model M_b fits the data. Again, we point out that Model M_0 is a special case of M_b, so the result is not surprising.

The final goodness of fit test is for Model M_t (test 6). In Fig. 13, the test statistic value of 59.465 (54 df) is not significant ($P = 0.28338$).

An alternate way of thinking of these tests is as follows. The test of Model M_0 vs. M_h tests whether there is heterogeneity of capture probabilities while the goodness of fit test of Model M_h tests whether there is any additional source of variation in capture probabilities due either to behavior or time or both. If we rejected Model M_0 in favor of M_h and the goodness of fit test to M_h indicates M_h fits the data, we could conclude Model M_h was appropriate.

Similarly, the tests of Model M_0 vs. M_b test for whether there is behavioral variation in capture probabilities, while the goodness of fit tests to Model M_b examine whether any other source of variation is operating.

The 7 tests have been simulated to study their size and power, and the results are presented in Appendix N. In general, all the tests have good size properties, i.e., they tend not to reject when the null hypothesis is true. However, most of the tests lack power, i.e., they do not always reject when they should. This is particularly true for tests of heterogeneity (M_0 vs. M_h and the goodness of fit test of M_h).

On the Need for an Objective Selection Procedure

The above discussion and Fig. 13 illustrate the general model selection approach. Conceptually, we examine the results of the 7 tests of assumptions, and on the basis of the results choose the appropriate model. In the example used (Fig. 13), the selection was not difficult. However, for other models, and for much real data we have seen, the judgemental selection of an appropriate estimator can be very difficult because (1) the tests are not independent, in particular if behavioral variation in capture probabilities is present this can strongly affect the test for time and heterogeneity effects; and (2) for real data all 3 sources of variability probably are present in varying degrees, so all we can hope to do is select the most appropriate model (the one that best describes the data). With respect to the first point, consider what happens if Model M_b is true. For example, assume 7 trapping occasions with M_b as the true model, and $p = 0.40$, $c = 0.20$. That is, prior to being captured animals have a 0.40 capture probability, but due to behavioral response (trap avoidance in this case) their capture probabilities drop to only 0.20 after first capture. The expected daily capture probability is given by

$$p_j = \frac{E(n_j)}{N}, \qquad j = 1, \ldots, 7.$$

These are constant only if the expected

daily captures are constant. But as trapping progresses the average daily capture probability decreases as more and more animals "shift" their capture probability from 0.40 to 0.20 as a result of trap response. A formula for p_j in this case is

$$p_j = \frac{(1-(1-p)^{j-1})\,c + (1-p)^{j-1}p^2}{1-(1-p)^j}, \qquad j = 2, \ldots, t$$

$(p_1 \equiv p)$. For the case above, we have

j	p_j
1	0.400
2	0.275
3	0.237
4	0.220
5	0.211
6	0.207
7	0.204

When Model M_b is true, the test for time variation in capture probabilities (Model M_0 vs. M_t) will tend to reject Model M_0 because the behavioral response does indeed "cause" time variation to be present also.

The same sort of argument shows that behavioral response "causes" heterogeneity on any given day. Consequently, when Model M_b is true, there is a strong tendency for all of the simple tests to reject Model M_0, making selection of the correct model difficult.

In the final analysis, selection of a model involves examining a point in a 7-dimensional space and trying to classify it into 1 of 8 categories. We have not been successful ourselves in providing a set of simple rules for this choice and believe that field biologists without rigorous statistical training might have great difficulty arriving at a proper choice.

An Objective Model Selection Procedure

The problem identified in the previous section falls in the realm of multivariate statistics, specifically classification of an unknown entity into one of several different known populations. In the classical version of this problem, one has a sample of measurements from each known population, and from those data a mathematical rule is constructed for classifying future cases based on their measurements. In our case, the "measurements" are the significance levels from the 7 tests. The "populations" are the 8 models. What is unknown is what model best fits any capture data at hand. Given this conceptualization, we chose to use discriminant function analyses to construct the classification function (see Cooley and Lohnes 1962). The objective of the discriminant analysis is to weight and linearly combine the significance levels of the 7 tests in some fashion so that the models are forced to be as statistically distinct as possible.

The model classification function was derived from simulated data where the true model is known. The capture probability structure of these simulated data is presented in Table 14. These parameter values were chosen because they are, in our opinion, representative of many real capture studies. For each population, 100 replications were generated, giving a total of 1,600 cases. The probability levels from each of the 7 tests were then used to construct the classification function. In addition to the 7 probabilities, 9 additional variables were constructed by taking all possible products between probability levels from tests 1–3 and tests 4–6, i.e. by taking the 9 pairwise products of the probabilities from the 3 specific tests against Model M_0 and the 3 goodness of fit tests. This provides a total of 16 variables that were transformed by $X = \log(P + 0.01)$. That transformation tends to weight the smaller probability levels more heavily than the larger values close to 1. The classification rule resulting from that series of transformations was one of many rules examined; this particular rule was chosen because it had the best performance.

The transformed variables were then used to compute a classification function using the SPSS discriminant procedure (Nie et al. 1975) . The classification func-

TABLE 14.—CAPTURE PROBABILITY STRUCTURE FOR THE SIMULATED DATA USED TO GENERATE THE CLASSIFICATION FUNCTION. ONE HUNDRED REPLICATIONS OF EACH TYPE WERE GENERATED, GIVING A TOTAL SAMPLE SIZE OF 1,600

Model	Probability Structure
	LARGE POPULATION (N = 400, t = 7)
M_o	$p = 0.30$.
M_h	$p_i = 0.15, i = 1{,}100; p_i = 0.25, i = 101{,}200; p_i = 0.35, i = 201{,}300; p_i = 0.45, i = 301{,}400$.
M_b	$p = 0.2; c = 0.5$.
M_{bh}	$p_i = 0.15, c_i = 0.375, i = 1{,}100; p_i = 0.25, c_i = 0.625, i = 101{,}300; p_i = 0.35, c_i = 0.875, i = 301{,}400$.
M_t	$p_1 = 0.15; p_2 = 0.45; p_3 = 0.35; p_4 = 0.25; p_5 = 0.30; p_6 = 0.20; p_7 = 0.40$.
M_{th}	$p_{ij} = p_i \cdot p_j; p_i = 0.45, i = 1{,}100; p_i = 0.55, i = 101{,}200; p_i = 0.65, i = 201{,}300; p_i = 0.75, i = 301{,}400; p_1 = 0.65; p_2 = 0.75; p_3 = 0.55; p_4 = 0.35; p_5 = 0.60; p_6 = 0.50; p_7 = 0.80$.
M_{tb}	$p_{ij} = p_j \cdot c; c = 2.50; p_1 = 0.05; p_2 = 0.35; p_3 = 0.25; p_4 = 0.15; p_5 = 0.20; p_6 = 0.10; p_7 = 0.30$.
M_{tbh}	$p_{ij} = p_i \cdot p_j \cdot c; p_i = 0.15, i = 1{,}100; p_i = 0.25, i = 101{,}200; p_i = 0.35, i = 201{,}300; p_i = 0.45, i = 301{,}400; c = 2.50; p_1 = 0.65; p_2 = 0.75; p_3 = 0.55; p_4 = 0.45; p_5 = 0.60; p_6 = 0.50; p_7 = 0.70$.
	SMALL POPULATION (N = 100, t = 5)
M_o	$p = 0.10$.
M_h	$p_i = 0.05, 1 = 1{,}25; p_i = 0.15, i = 26{,}50; p_i = 0.25, i = 51{,}75; p_i = 0.35, i = 76{,}100$.
M_b	$p = 0.40; c = 0.20$.
M_{bh}	$p_i = 0.15, c_i = 0.075, i = 1{,}33; p_i = 0.25, c_i = 0.125, i = 34{,}66; p_i = 0.35, c_i = 0.175, i = 67{,}100$.
M_t	$p_1 = 0.15; p_2 = 0.45; p_3 = 0.35; p_4 = 0.25; p_5 = 0.30$.
M_{th}	$p_{ij} = p_i \cdot p_j; p_i = 0.35, i = 1{,}25; p_i = 0.45, i = 26{,}50; p_i = 0.55, i = 51{,}75; p_i = 0.65, i = 76{,}100; p_1 = 0.65; p_2 = 0.75; p_3 = 0.55; p_4 = 0.35; p_5 = 0.60$.
M_{tb}	$p_{ij} = p_j \cdot c; c = 0.5; p_1 = 0.35; p_2 = 0.25; p_3 = 0.15; p_4 = 0.20; p_5 = 0.30$.
M_{tbh}	$p_{ij} = p_i \cdot p_j \cdot c; p_i = 0.15, i = 1{,}25; p_i = 0.25, i = 26{,}50; p_i = 0.35, i = 51{,}75; p_i = 0.45, i = 76{,}100; c = 0.75; p_1 = 0.65; p_2 = 0.75; p_3 = 0.55; p_4 = 0.45; p_5 = 0.60$.

tion consists of 8 sets of coefficients (1 set for each of the 8 models) that are used to form linear combinations of the 16 variables plus an intercept. The classification equations are derived from the pooled within-model covariance matrix and the centroids for the transformed probability levels. The i^{th} function has the form

$$C_i = c_{i0} + c_{i1} \cdot X_1 + c_{i2} \cdot X_2 + \cdots + c_{i16} \cdot X_{16}.$$

The 136 classification coefficients are

TABLE 15.—CLASSIFICATION FUNCTION COEFFICIENTS USED IN THE MODEL SELECTION PROCEDURE

Coefficient	M(O)	M(H)	M(B)	M(BH)	M(T)	M(TH)	M(TB)	M(TBH)
C0	−8.09154	−14.33446	−25.69392	−17.15981	−31.90722	−23.20432	−26.11839	−15.07368
C1	2.44865	1.55835	2.41361	2.09104	4.85352	4.52842	4.58145	1.01685
C2	0.01513	0.76894	0.34328	−0.30063	4.19868	3.36253	−1.18146	−0.78482
C3	5.60627	7.61750	4.45155	2.65005	−3.83028	−4.33794	3.81010	1.97090
C4	2.31654	2.43157	−1.90418	1.85727	4.11214	3.84041	0.07484	5.00658
C5	0.39009	0.24251	2.57294	2.95937	−0.37644	−0.30277	−0.73170	−1.36695
C6	1.59847	1.50362	5.40943	0.82379	3.29315	2.76447	4.77294	0.91642
C7	−1.07536	−0.75833	−1.39342	−0.92010	−1.87794	0.47319	−1.46921	−0.87191
C8	−0.95870	−2.99247	−0.29576	0.83512	−1.39938	−1.28509	−1.44532	−2.06608
C9	−4.23372	−5.76230	−1.30645	−3.97856	−4.93318	−4.85310	−3.75607	−7.44977
C10	0.73672	2.16608	1.71422	−0.11956	1.32575	2.28866	1.74703	2.68257
C11	−0.90843	−0.51321	0.06944	−2.67855	−2.69316	−2.01668	−1.73071	1.08225
C12	1.54069	2.72345	−4.56388	−2.53049	−5.55330	−2.64194	0.53878	2.03048
C13	−5.49376	−6.30792	−2.38615	−2.14175	4.38634	−0.03381	−2.06993	−5.17029
C14	−3.30107	−2.40404	−5.13204	−2.38473	−3.79996	−4.33330	−4.76823	−2.41632
C15	−0.19891	−1.42895	−2.26381	−0.48135	0.35309	−2.42112	−1.92578	−0.09665
C16	−2.04687	−3.46579	−4.06512	−1.73548	−4.64956	−1.59132	−3.89432	−1.80314

TABLE 16.—CAPTURE PROBABILITY STRUCTURE USED TO SIMULATE GOOD, MEDIUM, AND POOR DATA TO STUDY THE OPERATING CHARACTERISTICS OF THE MODEL SELECTION PROCEDURE. A POPULATION OF 400 WAS TRAPPED 7 OCCASIONS FOR 200 CASES

Model	Probability Structure
	GOOD DATA (Average $p \approx 0.35$)
M_o	$p = 0.35$.
M_h	$p_i = 0.5, i = 1{,}100; p_i = 0.4, i = 101{,}200; p_i = 0.3, i = 201{,}300; p_i = 0.2, i = 301{,}400$.
M_b	$p = 0.5; c = 0.2$.
M_{bh}	$p_i = 0.65, c_i = 0.10, i = 1{,}100; p_i = 0.55, c_i = 0.15, i = 101{,}200; p_i = 0.45, c_i = 0.10, i = 201{,}300; p_i = 0.35, c_i = 0.50, i = 301{,}400$.
M_t	$p_1 = 0.35, p_2 = 0.45, p_3 = 0.25, p_4 = 0.35, p_5 = 0.20, p_6 = 0.50, p_7 = 0.35$.
M_{th}	$p_{ij} = p_i \cdot p_j; p_i = 0.9, i = 1{,}100; p_i = 0.8, i = 101{,}200; p_i = 0.7, i = 201{,}300; p_i = 0.6, i = 301{,}400; p_1 = 0.5; p_2 = 0.6; p_3 = 0.4; p_4 = 0.5; p_5 = 0.35; p_6 = 0.65; p_7 = 0.5$.
M_{tb}	$p_{ij} = p_j \cdot c; c = 1.5; p_1 = 0.3; p_2 = 0.45; p_3 = 0.15; p_4 = 0.3; p_5 = 0.2; p_6 = 0.4; p_7 = 0.3$.
M_{tbh}	$p_{ij} = p_i \cdot p_j \cdot c; p_i = 0.5, i = 1{,}100; p_i = 0.4, i = 101{,}200; p_i = 0.3, i = 201{,}300; p_i = 0.2, i = 301{,}400; c = 1.5; p_1 = 0.7; p_2 = 0.7; p_3 = 0.4; p_4 = 0.4; p_5 = 0.7; p_6 = 0.9; p_7 = 0.87$.
	MEDIUM DATA (Average $p \approx 0.20$)
M_o	$p = 0.20$.
M_h	$p_i = 0.3, i = 1{,}100; p_i = 0.25, i = 101{,}200; p_i = 0.15, i = 201{,}300; p_i = 0.10, i = 301{,}400$.
M_b	$p = 0.15; c = 0.03$.
M_{bh}	$p_i = 0.35, c_i = 0.05, i = 1{,}100; p_i = 0.4, c_i = 0.1, i = 101{,}200; p_i = 0.2, c_i = 0.05, i = 201{,}300; p_i = 0.25, c_i = 0.2, i = 301{,}400$.
M_t	$p_1 = 0.02; p_2 = 0.35; p_3 = 0.05; p_4 = 0.2; p_5 = 0.1; p_6 = 0.3; p_7 = 0.2$.
M_{th}	$p_{ij} = p_i \cdot p_j; p_i = 0.7, i = 1{,}100; p_i = 0.55, i = 101{,}200; p_i = 0.45, i = 201{,}300; p_i = 0.3, i = 301{,}400; p_1 = 0.4; p_2 = 0.55; p_3 = 0.25; p_4 = 0.4; p_5 = 0.3; p_6 = 0.5; p_7 = 0.4$.
M_{tb}	$p_{ij} = p_j \cdot c; c = 0.33; p_1 = 0.30; p_2 = 0.45; p_3 = 0.15; p_4 = 0.30; p_5 = 0.20; p_6 = 0.40; p_7 = 0.30$.
M_{tbh}	$p_{ij} = p_i \cdot p_j \cdot c; p_i = 0.3, i = 1{,}100; p_i = 0.25, i = 101{,}200; p_i = 0.15, i = 201{,}300; p_i = 0.10, i = 301{,}400; c = 2.00; p_1 = 0.50; p_2 = 0.65; p_3 = 0.35; p_4 = 0.50; p_5 = 0.40; p_6 = 0.70; p_7 = 0.50$.
	POOR DATA (Average $p \approx 0.05$)
M_o	$p = 0.05$.
M_h	$p_i = 0.09, i = 1{,}100; p_i = 0.05, i = 101{,}300; p_i = 0.10, i = 301{,}400$.
M_b	$p = 0.075; c = 0.01$.
M_{bh}	$p_i = 0.07, c_i = 0.03, i = 1{,}100; p_i = 0.03, c_i = 0.08, i = 101{,}300; p_i = 0.01, c_i = 0.09, i = 301{,}400$.
M_t	$p_1 = 0.05; p_2 = 0.01; p_3 = 0.09; p_4 = 0.05; p_5 = 0.07; p_6 = 0.03; p_7 = 0.05$.
M_{th}	$p_{ij} = p_i \cdot p_j; p_i = 0.20, i = 1{,}100; p_i = 0.25; i = 101{,}200; p_i = 0.05, i = 201{,}300; p_i = 0.10, i = 301{,}400; p_1 = 0.33; p_2 = 0.10; p_3 = 0.10; p_4 = 0.333; p_5 = 0.50; p_6 = 0.50; p_7 = 0.33$.
M_{tb}	$p_{ij} = p_j \cdot c; c = 0.10; p_1 = 0.10; p_2 = 0.15; p_3 = 0.05; p_4 = 0.10; p_5 = 0.19; p_6 = 0.01; p_7 = 0.10$.
M_{tbh}	$p_{ij} = p_i \cdot p_j \cdot c; p_i = 0.20, i = 1{,}100; p_i = 0.25, i = 101{,}200; p_i = 0.05, i = 201{,}300; p_i = 0.10, i = 301{,}400; c = 2.00; p_1 = 0.20; p_2 = 0.15; p_3 = 0.20; p_4 = 0.05; p_5 = 0.05; p_6 = 0.15; p_7 = 0.37$.

given in Table 15. The 8 C_i values are then standardized over the interval 0 to 1 to give the model selection criteria. These standardized "model selection criteria" are printed by program CAPTURE right after the 7 tests (cf. Fig. 13).

Given the classification functions, an evaluation of their usefulness must be made. Again, data are required where the underlying models are known; this is only possible with simulated data. Hence, we simulated 3 additional data sets: 200 replications each of good, medium, and poor data for each model. In all cases, we used $N = 400$ and $t = 7$. The good data had an approximate overall average capture probability $\bar{p}$ of 0.35, medium $\bar{p}$ was approximately 0.20, and poor $\bar{p}$ was approximately 0.05. The capture probability structure of these data is given in Table 16.

We analyzed each data set and selected a model based on the classification function described above. This allows an evaluation of the selection procedure, i.e., how often a wrong model is chosen, and which models tend to be confused. These results are given in Tables 17–19

TABLE 17.—PERFORMANCE OF THE MODEL SELECTION PROCEDURE WITH GOOD DATA. THE TRUE MODEL FROM WHICH THE DATA WERE GENERATED IS AT THE TOP, AND THE MODEL SELECTED BY THE CLASSIFICATION FUNCTION ON THE SIDE. THE ENTRIES ARE THE PERCENTAGE OF TIMES THE MODEL WAS SELECTED, THE AVE($\hat{N}$), AND THE 95% CONFIDENCE INTERVAL COVERAGE

Model selected	Data generated from model								All data
	M_o	M_h	M_b	M_{bh}	M_t	M_{th}	M_{tb}	M_{tbh}	
M_o	70.0	2.5	0.0	0.0	0.0	0.0	0.0	0.0	9.1
	399.7	382.2	–	–	–	–	–	–	399.1
	0.957	0.200	–	–	–	–	–	–	0.931
M_h	21.5	93.5	0.0	0.0	0.0	0.0	0.0	0.0	14.4
	414.0	427.4	–	–	–	–	–	–	424.9
	0.349	0.059	–	–	–	–	–	–	0.113
M_b	0.0	0.0	90.0	0.0	0.0	0.0	0.0	0.0	11.3
	–	–	399.7	–	–	–	–	–	399.7
	–	–	0.922	–	–	–	–	–	0.922
M_{bh}	6.5	1.5	0.5	93.0	0.0	0.0	0.0	0.0	12.7
	394.2	396.3	399.0	396.6	–	–	–	–	396.4
	0.692	1.00	1.00	0.575	–	–	–	–	0.591
M_t	1.0	0.0	0.0	0.0	82.0	58.5	2.0	0.0	17.9
	404.0	–	–	–	399.8	396.3	372.0	–	398.0
	1.00	–	–	–	0.951	0.829	0.0	–	0.888
M_{th}	0.0	1.0	0.0	0.0	6.5	39.5	0.5	0.0	5.9
	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–
M_{tb}	0.0	0.5	9.5	0.0	11.0	2.0	95.0	37.0	19.4
	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–
M_{tbh}	1.0	1.0	0.0	7.0	0.5	0.0	2.5	63.0	9.4
	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–
Mean population estimate	402.5	425.8	399.7	396.6	399.8	396.3	372.0	–	
Coverage	0.808	0.077	0.923	0.575	0.951	0.829	0.0	–	

for the 3 types of data. However, our interest in these simulated data does not stop with the results of the model selection. Even more informative is how well the estimator performed when the wrong model was selected (we know that estimator performance usually is acceptable when the correct model is used). Information on the Ave($\hat{N}$) and confidence interval coverage is also given in Tables 17–19 for all estimators. The tables also provide summaries by model selection and by data type.

Those tables again emphasize the importance of high capture probabilities. The estimates and coverage coefficients drastically decline in usefulness as the capture probabilities decline. In addition, the correct model is much less likely to be selected for poor data. Note that in Table 19 (the poor data case), Model M_o is selected 34.7 percent of the time (whereas it is the true model only 12.5% of the time), and that Ave($\hat{N}_o$) is 745.5.

For good data, such a discrimination procedure will do quite well in selecting the appropriate model. We suggest that users knowledgeable in statistics and experienced in analysis of capture data may sometimes be able to render an improved judgement about the approrpriate model for real data. However, the field biologist should probably accept the model recommended by the model selection algorithm. The primary reason for this is the problems with dependence among tests

TABLE 18.—PERFORMANCE OF THE MODEL SELECTION PROCEDURE WITH MEDIUM DATA. THE TRUE MODEL FROM WHICH THE DATA WERE GENERATED IS AT THE TOP, AND THE MODEL SELECTED BY THE CLASSIFICATION FUNCTION ON THE SIDE. THE ENTRIES ARE THE PERCENTAGE OF TIMES THE MODEL WAS SELECTED, THE AVE($\hat{N}$), AND THE 95% CONFIDENCE INTERVAL COVERAGE

Model selected	Data generated from model								All data
	M_o	M_h	M_b	M_{bh}	M_t	M_{th}	M_{tb}	M_{tbh}	
M_o	68.0	28.0	0.0	0.0	0.0	0.0	0.0	0.5	12.1
	400.8	361.1	–	–	–	–	–	234.0	388.4
	0.971	0.089	–	–	–	–	–	0.000	0.710
M_h	18.5	66.5	0.5	0.0	0.0	0.0	0.0	0.0	10.7
	484.7	449.7	890.4	–	–	–	–	–	459.9
	0.000	0.399	0.000	–	–	–	–	–	0.310
M_b	1.0	0.0	81.0	58.0	0.0	0.0	0.0	1.5	17.7
	427.0	–	397.2	388.4	–	–	–	783.0	397.9
	0.500	–	0.951	0.681	–	–	–	1.000	0.837
M_{bh}	7.5	3.0	9.0	32.0	0.0	0.0	0.0	3.5	6.9
	409.3	382.0	471.8	390.1	–	–	–	467.1	410.6
	0.667	0.333	0.949	0.688	–	–	–	1.000	0.727
M_t	1.0	0.0	0.0	0.0	57.5	67.5	0.0	1.5	15.9
	405.0	–	–	–	400.3	377.9	–	251.3	386.7
	1.000	–	–	–	0.974	0.593	–	0.000	0.761
M_{th}	3.0	2.0	0.0	0.0	10.5	25.5	0.0	11.5	6.6
	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–
M_{tb}	0.0	0.0	8.5	8.5	31.5	6.5	98.0	65.0	27.2
	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–
M_{tbh}	1.0	0.5	1.0	1.5	0.5	0.5	2.0	16.5	2.9
	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–
Mean population estimate	418.0	422.2	407.4	389.0	400.3	377.9	–	471.9	
Coverage	0.755	0.308	0.945	0.683	0.974	0.593	–	0.714	

and the fact that with real data it is highly unlikely that any of these 8 models will be exactly "true."

Estimation in Alternative Models

When the model selection procedure described above has classified a model as the best one for a given set of data, 2 problems may still arise. First, the model may not have an associated estimation procedure for estimating N. Second, the model with the largest selection value (always 1.0) may not really fit the data, even though it is the "best" model available. In the first case, the investigator should scan for a large selection value (say 0.90) corresponding to a model having an estimator. If such a model is found, and if the relevant tests of model assumptions suggest that the goodness of fit of the model is adequate, the corresponding estimator can be used with some confidence. We would caution, however, against using models with selection values less than 0.75 to produce estimates of N, especially if there is a poor fit of the model to the data. In the second case, none of the models with corresponding estimation procedures seems to fit the data well; where fit is judged by the model selection procedure and inspection of individual tests, then we can see no justification for granting statistical validity to any calculated population estimate.

TABLE 19.—PERFORMANCE OF THE MODEL SELECTION PROCEDURE WITH POOR DATA. THE TRUE MODEL FROM WHICH THE DATA WERE GENERATED IS AT THE TOP, AND THE MODEL SELECTED BY THE CLASSIFICATION FUNCTION ON THE SIDE. THE ENTRIES ARE THE PERCENTAGE OF TIMES THE MODEL WAS SELECTED, THE AVE($\hat{N}$), AND THE 95% CONFIDENCE INTERVAL COVERAGE

Model selected	Data generated from model								All data
	M_o	M_h	M_b	M_{bh}	M_t	M_{th}	M_{tb}	M_{tbh}	
M_o	79.0	74.0	39.5	83.0	0.0	1.0	0.0	1.0	34.7
	417.5	316.3	3,229.7	270.2	–	342.5	–	146.5	745.5
	0.949	0.547	0.557	0.386	–	0.500	–	0.000	0.615
M_h	13.0	18.5	4.5	11.0	1.0	9.0	0.0	5.0	7.8
	377.7	336.5	612.2	255.6	374.6	347.1	–	210.5	342.8
	0.923	0.514	0.000	0.000	1.000	0.611	–	0.000	0.452
M_b	0.5	1.0	14.0	0.0	0.0	0.0	0.0	0.0	1.9
	158.0	156.0	272.8	–	–	–	–	–	261.6
	0.000	0.000	0.250	–	–	–	–	–	0.226
M_{bh}	3.0	4.0	33.0	2.0	0.0	0.0	0.0	0.0	5.3
	174.7	172.1	359.5	123.3	–	–	–	–	317.2
	0.000	0.000	0.924	0.000	–	–	–	–	0.726
M_t	1.0	0.5	0.0	1.5	83.5	24.5	2.0	68.0	22.6
	366.5	246.0	–	216.7	406.7	325.9	2,983.0	164.5	331.0
	1.000	0.000	–	0.000	0.940	0.653	0.000	0.059	0.550
M_{th}	1.5	0.5	0.5	1.5	5.0	5.0	0.0	0.5	1.8
	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–
M_{tb}	0.5	0.0	4.5	0.0	1.0	0.0	92.5	0.0	12.3
	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–
M_{tbh}	1.5	1.5	4.0	1.0	9.5	60.5	5.5	25.5	13.6
	–	–	–	–	–	–	–	–	–
	–	–	–	–	–	–	–	–	–
Mean population estimate	402.7	312.2	1,604.5	264.7	406.3	331.9	2,983.0	167.4	
Coverage	0.912	0.510	0.615	0.328	0.941	0.638	0.000	0.054	

Additional Examples of Model Selection

As part of a population ecology study on salt marsh rodents, Coulombe (1965, unpublished master's thesis, University of California, Los Angeles, California), conducted a livetrapping study on an outbreak of feral house mice *Mus musculus* in a salt marsh in mid-December 1962, at Ballana Creek, Los Angeles County, California. A square 10 × 10 grid was used with traps spaced 3 m apart and trapping was done twice daily, morning and evening, for 5 days. Thus there are 10 trapping occasions, but we can expect time variation in capture probabilities between morning and night occasions. The entire data set of 173 distinct individuals captured included young and adult, and male and female. Thus, we might also expect some heterogeneity of capture probabilities.

From the model selection procedure (Fig. 14), there is clear evidence of time

→

FIG. 14. Example of the model selection procedure applied to Coulombe's (unpublished thesis) full data set. Appropriate model probably is M_{th}. No estimator results from the model.

```
OCCASION          J=      1    2    3    4    5    6    7    8    9   10
ANIMALS CAUGHT  N(J)=    68   61   62   52   74   41   76   36   76   39
TOTAL CAUGHT    M(J)=     0   68  102  128  140  156  159  171  171  173  173
NEWLY CAUGHT    U(J)=    68   34   26   12   16    3   12    0    2    0
FREQUENCIES     F(J)=     2   64   40   31   16   13    5    1    0    1


1. TEST FOR HETEROGENEITY OF TRAPPING PROBABILITIES IN POPULATION.
   NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(H)

        CHI-SQUARE VALUE =    48.576   DEGREES OF FREEDOM =   6   PROBABILITY OF LARGER VALUE =  0.00000


2. TEST FOR BEHAVIORAL RESPONSE AFTER INITIAL CAPTURE.
   NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(B)

        CHI-SQUARE VALUE =     1.848   DEGREES OF FREEDOM =   1   PROBABILITY OF LARGER VALUE =   .17400


3. TEST FOR TIME SPECIFIC VARIATION IN TRAPPING PROBABILITIES.
   NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(T)

        CHI-SQUARE VALUE =    62.246   DEGREES OF FREEDOM =   9   PROBABILITY OF LARGER VALUE =  0.00000

4. GOODNESS OF FIT TEST OF MODEL  M(H)
   NULL HYPOTHESIS OF MODEL  M(H) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(H)

        CHI-SQUARE VALUE =    57.151   DEGREES OF FREEDOM =   9   PROBABILITY OF LARGER VALUE =  0.00000

        TEST OF MODEL M(H) BY FREQUENCY OF CAPTURE
        (FREQUENCIES LESS THAN 2T ARE NOT CALCULATED.)

        NUMBER OF CAPTURES  CHI-SQUARE  D.F.  PROBABILITY
        ---------------------------------------------------

               2               14.027     9      .12136
               3                6.857     9      .65199
               4               27.387     9      .00121

5. GOODNESS OF FIT TEST OF MODEL  M(B)
   NULL HYPOTHESIS OF MODEL  M(B) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(B)

        CHI-SQUARE VALUE =    68.087   DEGREES OF FREEDOM =  15   PROBABILITY OF LARGER VALUE =  0.00000

     5A. CONTRIBUTION OF TEST OF HOMOGENEITY OF FIRST CAPTURE PROBABILITY ACROSS TIME

        CHI-SQUARE VALUE =    24.780   DEGREES OF FREEDOM =   7   PROBABILITY OF LARGER VALUE =   .00083

     5B. CONTRIBUTION OF TEST OF HOMOGENEITY OF RECAPTURE PROBABILITIES ACROSS TIME

        CHI-SQUARE VALUE =    43.307   DEGREES OF FREEDOM =   8   PROBABILITY OF LARGER VALUE =   .00000


6. GOODNESS OF FIT TEST OF MODEL  M(T)
   NULL HYPOTHESIS OF MODEL  M(T) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(T)

        CHI-SQUARE VALUE =   162.329   DEGREES OF FREEDOM = 125   PROBABILITY OF LARGER VALUE =   .01386


7. TEST FOR BEHAVIORAL RESPONSE IN PRESENCE OF HETEROGENEITY.
   NULL HYPOTHESIS OF MODEL  M(H)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(BH)

        CHI-SQUARE VALUE =    31.439   DEGREES OF FREEDOM =  24   PROBABILITY OF LARGER VALUE =   .14153


MODEL SELECTION CRITERIA.  MODEL SELECTED HAS MAXIMUM VALUE.

MODEL       M(O)       M(H)       M(B)       M(BH)      M(T)       M(TH)      M(TB)      M(TBH)
CRITERIA     .15        .24        .03        .22        .26       1.00       0.00        .66
```

and heterogeneity variation, but little indication of behavioral variation. Note also that none of the simple models (M_h, M_b, M_t) fit the data. The model selection criteria strongly suggest Model M_{th} as appropriate, but there is no estimator associated with this model. To obtain an estimator, we can look for the next most likely model, which is M_{tbh} (selection criteria = 0.66); however, there is no estimator for that model either. In fact, to get a model allowing estimation we would have to use either M_t, M_h, or M_{bh}. Unfortunately, the value of the selection criteria corresponding to those models is too low to allow legitimate choice of one of the models for estimation purposes. Therefore, the investigator must realize that none of the models can be used to estimate population size with the data in the present form.

One alternative is to simply take the total number seen (M_{t+1}) as the best estimate. Given good data, M_{t+1} will be nearly equal to N. Also with good data one will tend to reject the simpler models. The real measure, however, of good data comes from such things as t, n./M_{t+1}, the pattern of the new captures (the u_j's) and the apparent average capture probability (average $\hat{p}$ is 0.3 here using $\hat{N} = 173 = M_{11}$). In good data, very few new animals will be caught by the last few occasions; in Coulombe's data, only 2 new animals were caught after the seventh occasion. Similarly, the capture frequency data should show many animals caught 2, 3, 4, or more times and there should not be a strong spike at f_1 (captured once only). By all these measures, Coulombe's data suggest that almost all the population was caught. Thus, here we would suggest taking 173 as the estimated population size.

Under those circumstances, we would also expect the point estimate of N from all 5 possible estimators to be in close agreement with the value of 173. In fact the results for Coulombe's full data set are:

Estimator	Value	Standard error
$\hat{N}_0$	176	1.8
$\hat{N}_t$	175	1.6
$\hat{N}_b$	174	1.6
$\hat{N}_h$	175	1.8
$\hat{N}_{bh}$	173	0.2

These are all very similar and precise. Note that in general it is not appropriate to compute all estimates. We recommend it only when there is strong evidence all animals were captured, in which case it provides an additional check.

As a further example, Fig. 15 shows the model selection results using only the morning capture data from Coulombe's (unpublished thesis) study (i.e., pretending the evening captures never occurred). Presumably, this would eliminate most of the time variation, but not heterogeneity. The selection criteria verify this conjecture.

Model	Model selection criteria
M_0	0.99
M_h	1.00
M_b	0.58
M_{bh}	0.74
M_t	0.00
M_{th}	0.46
M_{tb}	0.53
M_{tbh}	0.80

When the criteria value for 2 or more models exceed 0.95, the program does not just suggest 1, but names the 2 models that have the highest criteria. Thus, in this case the choice between models M_0 and M_h is not clear cut. As a standard operating procedure, we rec-

→

FIG. 15. Example of the model selection procedure applied to Coulombe's (unpublished thesis) morning trapping occasions (Occasions 1, 3, 5, 7, 9). Appropriate model probably is M_h or M_0. Suggested estimator is jackknife.

```
OCCASION              J=      1     2     3     4     5
ANIMALS CAUGHT     N(J)=     68    62    74    76    76
TOTAL CAUGHT       M(J)=      0    68   103   127   153   160
NEWLY CAUGHT       U(J)=     68    35    24    26     7
FREQUENCIES        F(J)=     45    59    36    15     5


1. TEST FOR HETEROGENEITY OF TRAPPING PROBABILITIES IN POPULATION.
   NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(H)

        CHI-SQUARE VALUE =     5.051   DEGREES OF FREEDOM =   3   PROBABILITY OF LARGER VALUE =   .16813


2. TEST FOR BEHAVIORAL RESPONSE AFTER INITIAL CAPTURE.
   NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(B)

        CHI-SQUARE VALUE =     2.271   DEGREES OF FREEDOM =   1   PROBABILITY OF LARGER VALUE =   .13184


3. TEST FOR TIME SPECIFIC VARIATION IN TRAPPING PROBABILITIES.
   NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(T)

        CHI-SQUARE VALUE =     3.667   DEGREES OF FREEDOM =   4   PROBABILITY OF LARGER VALUE =   .45295

4. GOODNESS OF FIT TEST OF MODEL  M(H)
   NULL HYPOTHESIS OF MODEL  M(H) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(H)

        CHI-SQUARE VALUE =     3.674   DEGREES OF FREEDOM =   4   PROBABILITY OF LARGER VALUE =   .45191

        TEST OF MODEL M(H) BY FREQUENCY OF CAPTURE
        (FREQUENCIES LESS THAN 2T ARE NOT CALCULATED.)

        NUMBER OF CAPTURES  CHI-SQUARE  D.F.  PROBABILITY
        ---------------------------------------------

                1                3.333    4       .50367
                2                4.362    4       .35928
                3                3.074    4       .54551
                4                2.667    4       .61506
 5. GOODNESS OF FIT TEST OF MODEL  M(B)
    NULL HYPOTHESIS OF MODEL  M(B) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(B)

         CHI-SQUARE VALUE =    10.240   DEGREES OF FREEDOM =   6   PROBABILITY OF LARGER VALUE =   .11492

     5A. CONTRIBUTION OF TEST OF HOMOGENEITY OF FIRST CAPTURE PROBABILITY ACROSS TIME

         CHI-SQUARE VALUE =     7.735   DEGREES OF FREEDOM =   3   PROBABILITY OF LARGER VALUE =   .05183

     5B. CONTRIBUTION OF TEST OF HOMOGENEITY OF RECAPTURE PROBABILITIES ACROSS TIME

         CHI-SQUARE VALUE =     2.505   DEGREES OF FREEDOM =   3   PROBABILITY OF LARGER VALUE =   .47440


 6. GOODNESS OF FIT TEST OF MODEL  M(T)
    NULL HYPOTHESIS OF MODEL  M(T) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(T)

         CHI-SQUARE VALUE =  115.230   DEGREES OF FREEDOM = 101   PROBABILITY OF LARGER VALUE =   .15766


 7. TEST FOR BEHAVIORAL RESPONSE IN PRESENCE OF HETEROGENEITY.
    NULL HYPOTHESIS OF MODEL  M(H)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(BH)

         CHI-SQUARE VALUE =   16.497   DEGREES OF FREEDOM =  10   PROBABILITY OF LARGER VALUE =   .08626


 MODEL SELECTION CRITERIA.  MODEL SELECTED HAS MAXIMUM VALUE.

 MODEL        M(O)       M(H)       M(B)       M(BH)      M(T)       M(TH)      M(TB)      M(TBH)
 CRITERIA      .99       1.00        .58        .74       0.00        .46        .53        .80
```

```
OVERALL TEST RESULTS --
Z-VALUE                                    -.403
PROBABILITY OF A SMALLER VALUE            .34362


TEST OF CLOSURE BY FREQUENCY OF CAPTURE.
(FREQUENCIES LESS THAN 10 ARE NOT COMPUTED.)

NUMBER OF CAPTURES  Z-VALUE  PROBABILITY
----------------------------------------

        2            -.660      .25477
        3            -.322      .37360
        4            -.066      .47356
        5             .694      .75607
        6           -1.085      .13906
```

FIG. 16. Example of the test for closure procedure with feral house mouse data from Coulombe (unpublished thesis).

ommend using the jackknife estimator in such cases because it is robust to heterogeneity and will tend to do well even if a competing model is true. For the morning data, the estimate of N from the jackknife estimator w as 194 ± 7.6, while that from the null estimator was 171 ± 4.1. Hence, both choices given by the model selection procedure provided similar estimates.

As a final check on whether time variation exists in the form of morning and evening differences, note that there were more captures in the mornings (occasions 1, 3, 5, 7, 9) than in the evening (occasions 2, 4, 6, 8, 10).

A Test for Closure

Throughout this monograph our philosophy has been that assumptions should be tested. One of the most critical assumptions behind this entire work is that of population closure. Although it is desirable to test closure, there are no truly suitable tests for this assumption. Closure is difficult to test for, because some types of variations in capture probabilities (especially behavioral) are difficult to distinguish from a failure of closure.

Burnham and Overton (pers. comm.) suggest a closure test based on Model M_h (cf. Appendix K). Fig. 16 gives an example produced by program CAPTURE of this closure test applied to Coulombe's (unpublished thesis) full data set. The test statistic value (z-value) is −0.403, and the significance level of the test is 0.34362. Thus, for Coulombe's data, closure is not rejected.

If individual capture frequencies are 10 or more, program CAPTURE also computes and prints a partitioned version of the closure test for those frequencies. For example, from Fig. 16, for animals caught twice the test value is −0.660 and is not significant. These partitioned test cases are of interest, but we emphasize the overall test is the one to use for the final judgement on closure.

Failure of closure means that during the study animals are either entering or leaving the population at risk of capture, or both. This could be caused, for example, by death, emigration, or the trap grid itself attracting animals from surrounding areas (especially likely in removal trapping). In any of those cases, the animals that enter or leave have zero capture probabilities during the time they are not part of the trapped population. The corresponding probability model has similarities to our behavioral Models M_b, M_{bh}, M_{tb}, and M_{tbh}. Indeed, as previously stated, behavioral responses are extremely difficult to untangle from true failures of closure. Our simulation results have shown the closure test rejects strongly when Model M_b is true, and is not a truly unbiased test whenever there is strong behavioral variation in capture probabilities. However, the test does not seem to be affected by heterogeneity or random time variations. We have not used other tests from the literature (e.g., Pollock et al. 1974) because, to our knowledge, those tests all are implicitly developed under the assumption that Model M_0 is the true model under the closure assumption. Because we feel Model M_0 probably is never acceptable, those tests will be even more untrustworthy than our current closure test.

The biologist is advised to look carefully at the closure test. If the test statistic is not significant, this tends to support the validity of the closure assumption. If that test rejects closure, before accepting that result, it is necessary to look at the other tests and the indicated model. If strong behavioral variation is indicated, the closure test is not valid.

Additional evidence regarding closure is obtained from a test of average captures per trap as a function of trap distance from the edge of the grid. That test is discussed in the section on DENSITY ESTIMATION; it tests for whether the grid attracts animals.

Finally, we mention that the closure test used here is oriented toward detecting breakdowns in closure only during the initial and final stages of the experiment. The test is not appropriate, for instance, for identifying situations in which animals emigrate during the middle of the study period and then immigrate back to the study area in the latter stages of an experiment.

DENSITY ESTIMATION

Introduction

The models discussed to this point involve only population size N as the parameter of interest. Often, interest may lie in population density, the number of animals per unit area (e.g., squirrels/hectare). One could naively take $\hat{N}$ divided by the area enclosed by the trapping grid as an estimate of density. That approach, however, leads to severe overestimation as a result of what has been called "edge effect," i.e., not all animals have their entire home range within the trapping grid, but may still be caught because some traps near the grid boundary are within their home range. Although biologists have recognized this problem for decades (Dice 1938, 1941; Stickel 1954), statisticians concerned with estimation of population abundance have tended to ignore or have failed to recognize the problem. This probably is due in part to the fact that abstract models for capture studies, such as ball-and-urn models, have no spatial component, hence do not include any concept of density as distinct from numbers of animals.

Three basic approaches are given in the biological literature to solve this problem. Two of them rely on Dice's (1938) concept of a boundary strip about the grid such that the effective trapping area is the grid area plus this boundary strip area. Dice assumed the boundary strip to be one-half the average diameter of the home range of the population being trapped. The first 2 approaches attempt to estimate this strip width, W, from the capture data. These 3 approaches are given below:

(1) Home range size is estimated from the locations of different captures for the same animal and is used to estimate the strip width W. A variety of approaches have been used; all are basically ad hoc and subject to numerous problems, e.g., results depend upon trap spacing and numbers of recaptures (Hayne 1949b, Stickel 1954, Tanaka 1972). The basic idea can be developed into an elaborate statistical estimation problem (Jennrich and Turner 1969), but as far as we know estimation of W based on movement data remains unsatisfactory.

(2) The parameter W is directly estimated based on data drawn from selected subgrids (MacLulich 1951, Hansson 1969, Seber 1973:51, Smith et al. 1975). As suggested by Burnham and Cushwa (pers. comm.) that idea can be developed into a procedure allowing joint estimation of D and W from data on 1 sufficiently large grid. We discuss this approach in detail below.

(3) The use of "assessment" lines is the most complex approach to density estimation. It involves designing the study to specifically estimate the effective trapping area as well as the size of the population at risk of capture. There are numerous variants on this approach (Smith et al. 1971, 1972, 1975, Swift and Steinhorst 1976, O'Farrell et al. 1977). We have not pursued this approach here because the proper data analysis depends upon the study design.

Of those 3 approaches, only the second seems to be formulated in a rigorous sta-

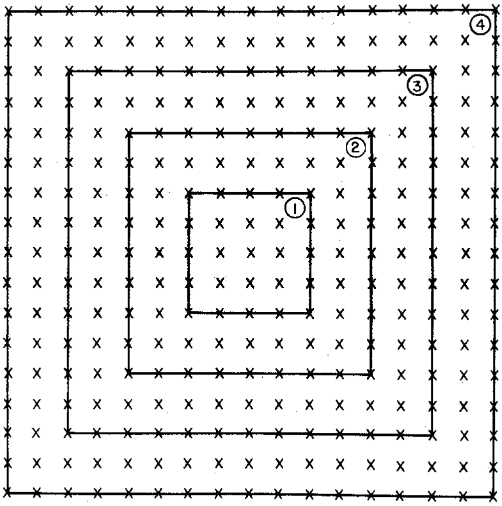


FIG. 17. A 17 × 17 trapping grid with 4 nested subgrids.

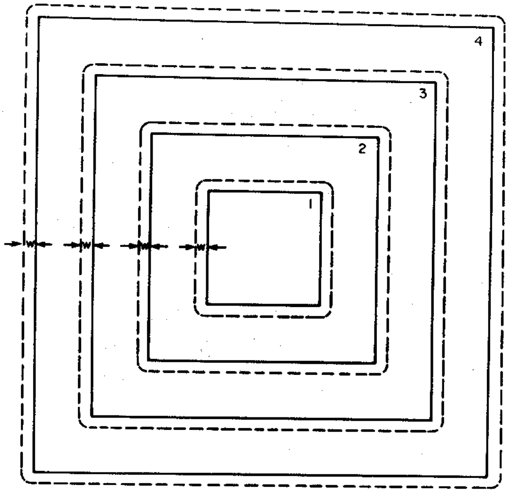


FIG. 18. A series of 4 nested subgrids with a boundary strip of width W around each subgrid. Note the corner of the boundary strip is a quarter circle of radius W.

tistical fashion without arbitrary assumptions. However, even the second approach cannot be made workable without some arbitrary ad hoc features.

Problem Formulation

We will illustrate the problem with some data from a livetrapping study of feral house mice *Mus musculus* in a southern California coastal salt marsh in December 1962 by Coulombe (unpublished thesis), that were also used to illustrate the model selection procedure. Only the basic specifications of the study will be needed here: 100 Sherman live traps were laid out in a 10 × 10 square grid. Morning and evening trapping sessions were conducted for 5 days and provided 10 trapping occasions, although only the morning sessions will be used in the calculations. Traps were 3 m apart and the area enclosed by the grid was 0.073 ha. We can delineate subgrids of smaller size by examining smaller groups of traps; in particular, we can conceptualize a series of nested subgrids (see for example Fig. 17).

The size of the population at risk of capture for each grid must be estimated. The choice of a population estimator should be based on the model selection procedures using the data for the entire grid. For $i = 1, 2, \ldots, k$, assume that a population estimate N_i has been made for the i^{th} grid by using data obtained only from that grid. The naive estimator of density of the i^{th} grid is given by

$$\hat{Y}_i = \hat{N}_i/A_i$$
$$i = 1, 2, \ldots, k,$$

TABLE 20.—THE INFORMATION NEEDED TO ESTIMATE DENSITY USING THE FOUR NESTED SUBGRIDS AND THE DATA FROM COULOMBE (UNPUBLISHED THESIS)

Grid			Area A_i (ha)	$\hat{N}_i$ (individuals)	$\hat{Y}_i$ $\left(\frac{\text{individuals}}{\text{ha}}\right)$	$\hat{S}e(\hat{Y}_i)$
No.	Size					
1	4 × 4	inner grid	0.0081	47	5770.9	607.8
2	6 × 6	middle inner grid	0.0225	107	4750.7	452.6
3	8 × 8	middle outer grid	0.0441	166	3768.5	243.4
4	10 × 10	entire grid	0.0729	194	2654.4	104.9

where A_i is the size (area) of the i^{th} trapping grid. The grid data are given in Table 20. Note that the standard error of $\hat{Y}_i$ is $\hat{S}e(\hat{N}_i)/A_i$.

The four $\hat{Y}_i$ values in Table 20 cannot reasonably be considered each an unbiased estimate of a single parameter. The bias is attributed to the phenomenon of edge effect, wherein the area used by individuals at risk of capture is not restricted to the area contained within the trapping grid; rather, there are areas outside the trapping grid that are part of the home ranges of individuals at risk of capture.

We assume that the population density is constant in the area of trapping. That is, there is not a trend in density across the grid. Under these conditions it is reasonable to think that there is an area of constant width about the trapping grid such that the actual area used by the population at risk of capture is the total area of the trapping grid plus that of the boundary strip. Let W equal the width of this strip. This is illustrated for the four grids in Fig. 18. The concept of a boundary strip goes back at least as far as Dice (1938). We need not interpret W, but from the inception of the boundary strip concept, biologists have considered that W is related to home range size (Dice 1938, 1941). In fact they arbitrarily assume W is one-half the maximum linear dimension of the average home range of the species.

Statistical Treatment

The approach to density estimation that we advocate for use with grid trapping is to formulate the problem as one of joint estimation of D and W, with the parameter N being a function of D and W. Then, by having 2 or more grids of different sizes, we can estimate those parameters with a weighted nonlinear least squares procedure.

In what follows, we assume that density (D) is expressed in individuals per unit area and strip width (W) in linear units, such as meters. Consider a layout of traps in the shape of a square or rectangle. That is, by connecting the outermost traps, a square or rectangle results. Let P be the measured perimeter of the grid. Let A be the area within this perimeter, and let A(W) equal the area obtained by adding a boundary strip of constant width W. Then the fundamental relationships is

$$A(W) = A + PW/c + \pi W^2/c,$$

where c is a conversion factor to express PW or W^2 in units of A. For example, to convert m^2 to ha, c = 1,000 m^2/ha.

Let there be k such grids identified. These may be subgrids of one grid, or may be physically separate study grids. Assume a constant density applies for each grid. Then, for grid i we would expect to have

$$N_i = D\,A_i(W) = D[A_i + P_iW/c + \pi\,W^2/c],$$

where N_i is the population at risk of capture with respect to the i^{th} grid of traps only. Conceptually, we should let the actual numbers of individuals at risk of capture on grid i be a random variable with expectation $DA_i(W)$. Then we would write $E(\hat{N}_i) = DA_i(W)$, and it is really $E(\hat{N}_i)$ we are estimating for the i^{th} grid. Letting i = 1, . . ., k we obtain structural equations relating the parameters D, W, and the induced parameters N_i. The area A_i, and the grid perimeter P_i must be known. Next we redefine the basic structural equations as

$$Y_i = \frac{N_i}{A_i} = D[1 + a_iW + b_iW^2] \qquad i = 1, 2, \ldots, k,$$

where

$$a_i = \frac{P_i}{A_ic} \text{ and } b_i = \frac{\pi}{A_ic}.$$

Assume that from the trapping data of each grid we have estimates of the N_i, expressed as $\hat{N}_i$, and estimates of their variances $Var(\hat{N}_i)$. Assume further that the $\hat{N}_i$ are good estimates in the sense that they have small bias, so we can write

$$\frac{\hat{N}_i}{A_i} = \hat{Y}_i = D[1 + a_iW + b_iW^2] + \epsilon_i, \qquad i = 1, 2, \ldots, k,$$

where ϵ_i is a random error with $E(\epsilon) = \underline{0}$, and covariance matrix $E(\underline{\epsilon}\ \underline{\epsilon}') = \underset{\sim}{\Sigma}$. The small bias of the $\hat{N}_i$ is a big assumption that relies heavily on previous sections. Note that $\hat{Y}_i = \hat{N}_i/A_i$ is what we would call the naive estimator of density, D, from the i^{th} trapping grid; clearly it can be significantly biased unless A_i is large relative to $(P_iW/c + \pi W^2/c)$, the area that is added by assuming a strip width of W. Note that for a sequence of k subgrids of increasing size, the most biased estimate of D is obtained from $\hat{Y}_1$ because A_1 is smallest relative to $(P_1W/c + \pi W^2/c)$. On the other hand, $\hat{Y}_k$ is the best estimate of D because A_k is largest relative to $(P_kW/c + \pi W^2/c)$. If we had a very large grid, say the state of Wyoming, the contribution of area due to the boundary strip W would be negligible, and we could ignore the problem. Because this is not the case, we use our biased estimates of Y_i to find an unbiased estimate of D.

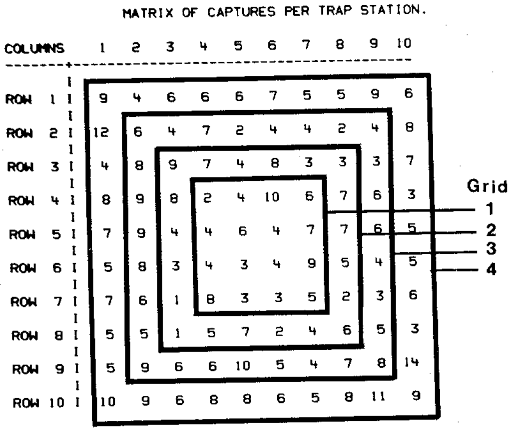
We assume $E(\underline{\epsilon}\ \underline{\epsilon}') = \underset{\sim}{\Sigma}$ because in general the naive density estimates, $\hat{Y}_i$, are not independent from one another if these subgrids derive from 1 overall grid. Hence, we must estimate the covariance matrix $\underset{\sim}{\Sigma}$ of the k estimates of $\hat{Y}_i$. A simple and intuitive procedure is to assume the correlation between $\hat{Y}_i$ and $\hat{Y}_j$ is just the proportion of overlapping area between grid i and grid j including their boundary strip. Then the covariance of $\hat{Y}_i$ and $\hat{Y}_j$ is $Se(\hat{Y}_i) \cdot Se(\hat{Y}_j) \cdot Corr(\hat{Y}_i, \hat{Y}_j)$. With the resulting covariance matrix, generalized nonlinear least squares can be performed to estimate density and strip width as $\hat{D}$ and $\hat{W}$, respectively. Note that $\underset{\sim}{\Sigma}$ is a function of W, because the amount of overlap between grid i and grid j is a function of the estimated strip width W.

In the most common situation, there is one grid such as our example in the California salt marsh, and one must choose a small number of subgrids to use in this approach. We recommend 4 or more nested grids. The main concern is to keep the inner grid large enough so that N_1 is well estimated. Other possibilities (not recommended) are to include subgrids consisting of halves and quarters of the total grid, or both.

The basic idea behind this procedure was introduced by MacLulich (1951), but until recently no method of estimating the variances of $\hat{D}$ and $\hat{W}$ had been given (Smith et al. 1975).

In Fig. 19, a complete analysis is presented for the California salt marsh feral mouse study. First note that we have computed the matrix of captures per trap station. That matrix includes the data for both the morning and evening trapping sessions, or for all 10 occasions. Multiple captures per trap cause some entries to be greater than 10. Visual study of this matrix (Fig. 19a) does not disclose any gross trends in mouse density across the grid. Fig. 19b shows the results of 3 chi-square goodness of fit tests, where the capture matrix is first collapsed by rows, then by columns, and finally by rings. Those tests generally tend to reject the null hypothesis of uniform probability of capture by rows or by columns. However, we are more concerned in identifying gross irregularities in mouse density, such as no captures in one corner of the grid, or a strong trend of decreasing probability of capture from right to left. The tests identify a problem of higher trap success at the grid edges. This is indicated by the test of rows, where a large portion of the chi-square value is contributed by the ninth and tenth rows, and by the ring test, where the outside ring of traps had much better success than expected, and the inner ring poorer success than expected. For large trapping grids, the problem could be corrected by not using the data for the outer ring of traps. This is very wasteful of data, however, and we will ignore the problem here.

The next 4 pages of output (Figs. 19c–f) consist of the population estimates for the whole grid and each of the 4 subgrids, computed with the jackknife estimator derived from Model M_h. That model was selected based on the output for the tests



```
           MATRIX OF CAPTURES PER TRAP STATION.

COLUMNS    1   2   3   4   5   6   7   8   9  10
-------+----------------------------------------
       I
ROW  1 I   9   4   6   6   6   7   5   5   9   6
       I
ROW  2 I  12   6   4   7   2   4   4   2   4   8
       I
ROW  3 I   4   8   9   7   4   8   3   3   3   7
       I
ROW  4 I   8   9   8   2   4  10   6   7   6   3
       I
ROW  5 I   7   9   4   4   6   4   7   7   6   5
       I
ROW  6 I   5   8   3   4   3   4   9   5   4   5
       I
ROW  7 I   7   6   1   8   3   3   5   2   3   6
       I
ROW  8 I   5   5   1   5   7   2   4   6   5   3
       I
ROW  9 I   5   9   6   6  10   5   4   7   8  14
       I
ROW 10 I  10   9   6   8   8   6   5   8  11   9
```

FIG. 19a. Nested subgrids used in the density estimation procedure with feral house mouse data from Coulombe (unpublished thesis). In the matrix, trap coordinates are rounded to the nearest whole integer. In goodness of fit tests, trap coordinates that are not integers and nonrectangular trapping grids will cause spurious results.

described in the section on model selection (Fig. 15).

The final page of output (Fig. 19g) gives the generalized least squares estimates of D and W. The columns headed NAIVE DENSITY, PERIMETER/AREA, and PI/AREA correspond to $\hat{Y}_i$, a_i, and b_i, respectively, and the COVARIANCE MATRIX corresponds to $\hat{\Sigma}$ for the initial value of W. The number of iterations required to estimate the parameters is given, and for this example it is 189. The maximum number allowed in program CAPTURE is 300. The approximate number of significant digits of $\hat{D}$ and $\hat{W}$ is also printed. A table listing the grid number, the naive density, and the predicted value is then given. This allows the investigator to get a feel for how well the model fits his data. Note that all the residuals in the example are negative; this is because they are all highly correlated, and hence we expect them to be of the same sign. The multiple correlation coefficient, R, is also printed to help assess the fit of the model. The value of R^2 is interpreted as the proportion of the variance in the data that is explained by the model ($0 \leq R^2 \leq 1$). The computed values of $\hat{D}$ and $\hat{W}$ are highly negatively correlated, indicating the

```
        CHI-SQUARE TEST OF UNIFORM DENSITY BY ROWS.

ROW                 1        2        3        4        5        6        7        8        9       10
OBSERVED           63       53       56       63       59       50       44       43       74       80
EXPECTED       58.500   58.500   58.500   58.500   58.500   58.500   58.500   58.500   58.500   58.500
CHI-SQUARE       .346     .517     .107     .346     .004    1.235    3.594    4.107    4.107    7.902

 TOTAL CHI-SQUARE =      22.26  WITH   9 DEGREES OF FREEDOM.  PROBABILITY OF LARGER VALUE =  .0081


        CHI-SQUARE TEST OF UNIFORM DENSITY BY COLUMNS.

COLUMN              1        2        3        4        5        6        7        8        9       10
OBSERVED           72       73       48       57       53       53       52       52       59       66
EXPECTED       58.500   58.500   58.500   58.500   58.500   58.500   58.500   58.500   58.500   58.500
CHI-SQUARE      3.115    3.594    1.885     .038     .517     .517     .722     .722     .004     .962

 TOTAL CHI-SQUARE =      12.08  WITH   9 DEGREES OF FREEDOM.  PROBABILITY OF LARGER VALUE =  .2090


        CHI-SQUARE TEST OF UNIFORM DENSITY BY RINGS (OUTER RING IS NUMBER 1).

RING                1        2        3        4        5
OBSERVED          247      160       96       65       17
EXPECTED      210.600  163.800  117.000   70.200   23.400
CHI-SQUARE      6.291     .088    3.769     .385    1.750

 TOTAL CHI-SQUARE =      12.28  WITH   4 DEGREES OF FREEDOM.  PROBABILITY OF LARGER VALUE =  .0154
```

FIG. 19b. Chi-square tests of uniform density with feral house mouse data from Coulombe (unpublished thesis).

```
NUMBER OF TRAPPING OCCASIONS WAS               5
NUMBER OF ANIMALS CAPTURED, M(T+1), WAS       33
TOTAL NUMBER OF CAPTURES, N., WAS             50

FREQUENCIES OF CAPTURE,F(I)
  I=   1   2   3   4   5
F(I)=  18  13  2   0   0

          COMPUTED JACKKNIFE COEFFICIENTS

          N(1)      N(2)      N(3)      N(4)      N(5)
      1  1.800     2.400     2.800     3.000     3.000
      2  1.000      .550      .050     -.250     -.250
      3  1.000     1.000     1.133     1.250     1.250
      4  1.000     1.000     1.000      .992      .992
      5  1.000     1.000     1.000     1.000     1.000


          THE RESULTS OF THE JACKKNIFE COMPUTATIONS

 I        N(I)   SE(I)     .95 CONF. LIMITS      TEST OF N(I+1) VS. N(I)
 0       33                                         CHI-SQUARE(1 D.F.)
 1       47.4     5.09      37.4      57.4          2.839
 2       52.4     7.57      37.5      67.2           .148
 3       53.3     9.51      34.7      72.0           .002
 4       53.3    10.62      32.4      74.1          0.000
 5       53.3    10.62      32.4      74.1          0.000

AVERAGE P-HAT =  .2128


INTERPOLATED POPULATION ESTIMATE IS          47 WITH STANDARD ERROR    4.9234

APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL       37 TO        57


HISTOGRAM OF F(I)

FREQUENCY     18   13    2    0    0
----------------------------------------
EACH * EQUALS    2 POINTS

       18      *
       16      *
       14      *    *
       12      *    *
       10      *    *
        8      *    *
        6      *    *
        4      *    *
        2      *    *    *
----------------------------------------
```

FIG. 19c. Example of population estimation with variable probability of capture by animal under Model M_h with feral house mouse data (Coulombe (unpublished thesis) from inner inner grid of Fig. 19a (X = 4–7, Y = 4–7).

```
NUMBER OF TRAPPING OCCASIONS WAS               5
NUMBER OF ANIMALS CAPTURED, M(T+1), WAS       67
TOTAL NUMBER OF CAPTURES, N., WAS            110

FREQUENCIES OF CAPTURE,F(I)
  I=   1   2   3   4   5
F(I)=  37  21  5   4   0

          COMPUTED JACKKNIFE COEFFICIENTS

          N(1)      N(2)      N(3)      N(4)      N(5)
      1  1.800     2.400     2.800     3.000     3.000
      2  1.000      .550      .050     -.250     -.250
      3  1.000     1.000     1.133     1.250     1.250
      4  1.000     1.000     1.000      .992      .992
      5  1.000     1.000     1.000     1.000     1.000


          THE RESULTS OF THE JACKKNIFE COMPUTATIONS

 I        N(I)   SE(I)     .95 CONF. LIMITS      TEST OF N(I+1) VS. N(I)
 0       67                                         CHI-SQUARE(1 D.F.)
 1       96.6     7.30      82.3     110.9         10.573
 2      109.4    10.91      88.0     130.7          2.231
 3      114.3    13.65      87.6     141.1           .789
 4      116.0    15.17      86.2     145.7          0.000
 5      116.0    15.17      86.2     145.7          0.000

AVERAGE P-HAT =  .2056


INTERPOLATED POPULATION ESTIMATE IS         107 WITH STANDARD ERROR   10.1827

APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL       86 TO       127


HISTOGRAM OF F(I)

FREQUENCY     37   21    5    4    0
----------------------------------------
EACH * EQUALS    4 POINTS

       36      *
       32      *
       28      *
       24      *
       20      *    *
       16      *    *
       12      *    *
        8      *    *
        4      *    *    *    *
----------------------------------------
```

FIG. 19d. Example of population estimation with variable probability of capture by animal under Model M_h with feral house mouse data (Coulombe unpublished thesis) from the middle inner grid of Fig. 19a (X = 3–8, Y = 3–8).

strong inverse relation between density and strip width. This negative correlation makes it difficult to estimate either parameter with a small standard error, because other combinations of $\hat{D}$ and $\hat{W}$ also result in almost as good a fit as the values selected. Finally, a test of whether $\hat{W}$ is significantly different from zero is printed. In that example, the highly significant difference of $\hat{W}$ from zero indicates that the estimate of D is much better than the naive estimate based on the actual grid area.

Discussion

The density estimation problem results because the grid is an artificial entity in the environment, and animals trapped use areas both inside and outside the grid. A practical problem illustrated in the example is that the grid can attract animals, thus "inducing" a higher density than would otherwise be found. The problem of the grid attracting animals can cause severe bias. If attraction occurs during trapping, then the assumption of population closure will be violated. With removal trapping, over a long enough time the problem is sure to develop. The problem may still occur in livetrapping studies. One possible test for this effect is the "ring" test given in Fig. 19b. Even with some approach like prebaiting to allow this attraction effect to stabilize before trapping, both methods (recapture movements, and direct estimation of W and D) will be invalid if the grid itself attracts animals. In that case, it appears assessment lines would have to be used.

If the above method produces a poor result, an alternative approach is to base an estimate of W on animal movements as determined from recapture locations.

```
NUMBER OF TRAPPING OCCASIONS WAS             5
NUMBER OF ANIMALS CAPTURED, M(T+1), WAS    116
TOTAL NUMBER OF CAPTURES, N., WAS          202

FREQUENCIES OF CAPTURE,F(I)
   I=   1   2   3   4   5
F(I)=  55  43  11   7   0

          COMPUTED JACKKNIFE COEFFICIENTS

          N(1)      N(2)      N(3)      N(4)      N(5)
     1   1.800     2.400     2.800     3.000     3.000
     2   1.000      .550      .050     -.250     -.250
     3   1.000     1.000     1.133     1.250     1.250
     4   1.000     1.000     1.000      .992      .992
     5   1.000     1.000     1.000     1.000     1.000


          THE RESULTS OF THE JACKKNIFE COMPUTATIONS

  I       N(I)    SE(I)     .95 CONF. LIMITS      TEST OF N(I+1) VS. N(I)
  0       116                                        CHI-SQUARE(1 D.F.)
  1       160.0     8.90     142.6     177.4          6.866
  2       173.7    13.20     147.8     199.5           .195
  3       175.6    16.64     143.0     208.2           .073
  4       174.9    18.62     138.4     211.4          0.000
  5       174.9    18.62     138.4     211.4          0.000

AVERAGE P-HAT =   .2434


INTERPOLATED POPULATION ESTIMATE IS        166 WITH STANDARD ERROR    10.7335

APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL       145 TO        188



HISTOGRAM OF F(I)

FREQUENCY     55   43   11    7    0
-----------------------------------------
EACH * EQUALS    6 POINTS

       54      *
       48      *
       42      *    *
       36      *    *
       30      *    *
       24      *    *
       18      *    *
       12      *    *    *
        6      *    *    *    *
-----------------------------------------
```

FIG. 19e. Example of population estimation with variable probability of capture by animal under Model M_h with feral house mouse data (Coulombe unpublished thesis) from the middle outer grid of Fig. 19a (X = 2–9, Y = 2–9).

```
NUMBER OF TRAPPING OCCASIONS WAS             5
NUMBER OF ANIMALS CAPTURED, M(T+1), WAS    160
TOTAL NUMBER OF CAPTURES, N., WAS          356

FREQUENCIES OF CAPTURE,F(I)
   I=   1   2   3   4   5
F(I)=  45  59  36  15   5

          COMPUTED JACKKNIFE COEFFICIENTS

          N(1)      N(2)      N(3)      N(4)      N(5)
     1   1.800     2.400     2.800     3.000     3.000
     2   1.000      .550      .050     -.250     -.250
     3   1.000     1.000     1.133     1.250     1.250
     4   1.000     1.000     1.000      .992      .992
     5   1.000     1.000     1.000     1.000     1.000


          THE RESULTS OF THE JACKKNIFE COMPUTATIONS

  I       N(I)    SE(I)     .95 CONF. LIMITS      TEST OF N(I+1) VS. N(I)
  0       160                                        CHI-SQUARE(1 D.F.)
  1       196.0     8.05     180.2     211.8           .007
  2       196.5    11.69     173.5     219.4          2.000
  3       189.8    15.15     160.1     219.4          2.847
  4       185.1    17.31     151.2     219.0          0.000
  5       185.1    17.31     151.2     219.0          0.000

AVERAGE P-HAT =   .3670


INTERPOLATED POPULATION ESTIMATE IS        194 WITH STANDARD ERROR     7.6455

APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL       178 TO        209



HISTOGRAM OF F(I)

FREQUENCY     45   59   36   15    5
-----------------------------------------
EACH * EQUALS    6 POINTS

       60           *
       54           *
       48           *
       42      *    *
       36      *    *    *
       30      *    *    *
       24      *    *    *
       18      *    *    *
       12      *    *    *    *
        6      *    *    *    *    *
-----------------------------------------
```

FIG. 19f. Example of population estimation with variable probability of capture by animal under Model M_h with feral house mouse data (Coulombe unpublished thesis) from the entire grid of Fig. 19a (X = 1–10, Y = 1–10).

Let $\tilde{W}$ be such an estimate, for example one-half of the average maximum distance between trapped locations for all animals captured at least twice. The standard error of $\tilde{W}$, $\widehat{\mathrm{Se}}(\tilde{W})$, can be computed from the data themselves. The estimator of density is then

$$\tilde{D} = \frac{\hat{N}}{A(\tilde{W})},$$

where $\hat{N}$ is based on all the data (i.e., the entire grid) and

$$A(\tilde{W}) \doteq A_g[1 + a\tilde{W} + b\tilde{W}^2],$$

for a and b as defined previously in this section. An estimator of the sampling variance of $\tilde{D}$ is given by

$$\widehat{\mathrm{Var}}(\tilde{D}) = \frac{\widehat{\mathrm{Var}}(\hat{N})}{[A(\tilde{W})]^2} + (\tilde{D})^2\left(\frac{a + 2b\tilde{W}}{1 + a\tilde{W} + b\tilde{W}^2}\right)^2 \widehat{\mathrm{Var}}(\tilde{W}),$$

and the standard error of $\tilde{D}$ is simply

$$\widehat{\mathrm{Se}}(\tilde{D}) = \sqrt{\widehat{\mathrm{Var}}(\tilde{D})}.$$

These variance formulas are valid whatever the technique for estimating W from recapture locations.

In the example of the feral house mice, $\tilde{W}$ is calculated as 3.63 m with $\widehat{\mathrm{Se}}(\tilde{W}) = 0.149$, and compares favorably with the estimated value of $\tilde{W} = 4.65$ from Fig. 19g.

The methods outlined in this section require large amounts of data to achieve satisfactory results. Both a large trapping grid and a large number of captures are required. A grid size of 9 × 9 probably can be considered the minimum; however, a larger grid such as 15 × 15 is much better. Good trapping success to achieve a large number of captures is necessary to provide a useful population

```
        STARTING VALUES FOR DENSITY ESTIMATION--
        NUMBER OF GRIDS                          4
        TRAP INTERVAL                         3.00
        UNITS CONVERSION                  10000.00
        INITIAL DENSITY ESTIMATE             .5089
        INITIAL STRIP WIDTH ESTIMATE      882.6797

   GRID NAIVE DENSITY  PERIMETER/AREA    PI/AREA          STARTING COVARIANCE MATRIX
     I      Y(I)               A(I)       B(I)

     1    5770.927         .4444444     .3879E-01      .369E+06
     2    4750.710         .2666667     .1396E-01      .273E+06 .205E+06
     3    3768.521         .1904762     .7124E-02      .146E+06 .109E+06 .592E+05
     4    2654.402         .1481481     .4309E-02      .625E+05 .468E+05 .254E+05 .110E+05


RESULTS OF ITERATIONS
   FUNCTION EVALUATIONS REQUIRED                          189
   ESTIMATED SIGNIFICANT DIGITS OF PARAMETER VALUES         8


        FITTED MODEL COMPARED TO THE DATA
        GRID(I)      Y(I)       F(I)
            1     5770.927    5505.869
            2     4750.710    3583.134
            3     3768.521    2875.007
            4     2654.402    2511.652

MULTIPLE CORRELATION COFFICIENT IS  .93181

ESTIMATED DENSITY=       1408.934  172.1260   = ITS STANDARD ERROR
ESTIMATED STRIP WIDTH=      4.653    1.0576   = ITS STANDARD ERROR
CORRELATION OF ESTIMATORS     -.9471

TEST OF ESTIMATED STRIP WIDTH GREATER THAN ZERO.
Z-VALUE =  4.3998  PROBABILITY OF LARGER VALUE =    .0000


          FINAL COVARIANCE MATRIX

        .3694E+06
        .1560E+06 .2048E+06
        .5398E+05 .7085E+05 .5924E+05
        .1821E+05 .2127E+05 .1779E+05 .1100E+05
```

FIG. 19g. Example of joint estimation of density and boundary strip width with feral house mouse data from Coulombe (unpublished thesis).

estimate for each grid. This becomes a problem especially with the smaller grids, where only a fraction of the data are used to estimate the population. Besides requiring a good population estimate, a good variance estimate for $\hat{N}$ is required because that quantity is used in constructing the estimate of the weighting matrix $\hat{\Sigma}^{-1}$. Poor data result in a poor weighting matrix, which then results in poor estimates of D and W.

Possible methods to increase the amount of data are (1) to place traps close together to increase recaptures, (2) to place grids in a uniform habitat so that discontinuities in density do not occur, and (3) to increase the number of traps. To repeat, the method requires large amounts of data. A carefully designed study is required to obtain reliable values of $\hat{D}$ and $\hat{W}$, and only rarely can a typical capture–recapture study be made to yield reasonable estimates.

STUDY DESIGN

The objective of this monograph is to present methods for the thorough analysis of capture data when the target populations are closed except for known removals. However, proper planning, design, and field conduct of such studies is necessary to obtain meaningful data for analysis. Many factors must be considered when planning a capture–recapture or removal study to estimate animal abundance. We consider these as falling into 2 broad categories: statistical design and data recording considerations, and field procedures, although the distinction may sometimes seem a bit arbitrary. Examples of such considerations include (but are not limited to) the following:

Statistical considerations

number of capture occasions
time between occasions
size and shape of trapping grid
spacing of traps
number of traps at a point
numbering of traps

Field procedure considerations

live vs. removal captures
choice of trap type
choice of electrofishing gear
method of marking or tagging
method of recapture
use of bait
time of day to check traps
handling the animals

The purpose of this section is to present some suggestions and cautions regarding the aspects of study design. The general themes presented are to conduct the experiment so that (1) assumptions can be tested, (2) the closure assumption is met, (3) the simplest possible model is appropriate, and (4) the number of animals captured is maximized (including recaptures). We begin by mentioning livetrapping versus removal methods, then discuss closure, proceed through ways of eliminating variation due to time, behavior, and heterogeneity, and then discuss sample size considerations such as grid size and number of traps.

For additional discussion of design consideration in grid trapping the reader is referred to Overton and Davis (1969), Smith et al. (1969, 1971, 1975), Tanaka (1970), and Hansson (1974).

It is beyond the intended scope of this monograph to provide guidance on the operational aspects of capture studies, although they are important. For example, if the method of marking (or tagging) is such that marks are lost, then a basic assumption needed for meaningful results is violated. When making decisions about a field study, the scientist should consider the probable effects of the experimental design on assumptions necessary for data analysis. For additional discussion of field procedures the reader is referred to Davis (1956), Southwood (1966), and Taber and Cowan (1969).

Livetrapping Versus Removal Methods

As discussed in the previous sections, removal methods are a special case of livetrapping methods. That is, the removal estimators can be used on livetrapping data. Hence, we recommend that livetrapping methods should be used if possible because of the wider array of options available for the data analysis. A hazard of removal studies is that they disrupt the population, and as substantial animals are removed, immigration may occur; this violates closure.

For livetrapping studies, all possible precautions should be taken to prevent deaths of the animals while in the traps, e.g., shading the traps in summer, or avoiding periods of extreme cold. Similarly, it is assumed (implicitly) that the method of marking will not induce mortality. If substantial mortality cannot be avoided, then one must analyze the study as a removal experiment.

Closure

For the analysis methods presented here, the single most important assumption is closure. Closure is very difficult to test for, yet any violation of this assumption biases the tests and population estimators presented here.

Considerations to help assure closure include, for example, timing the trapping to avoid known migration times, and periods of recuitment (e.g., juveniles becoming trappable) or immigration. Also keep the duration of the experiment as short as possible. If it is necessary to study the population at such times, open population models should be used for data analysis (Seber 1973, Pollock 1975, Arnason and Baniuk 1977).

Grids are often thought to attract animals. For example, when kill traps are used, animals from the area around the grid will move onto the grid as local animals are removed (cf. Gentry et al. 1968, Smith et al. 1975:38). This violates closure, and the only good solution seems to be to keep the length of the study (in days) very short so that the study will end before significant immigration can occur.

Eliminating Variation Due to Time, Behavior, and Heterogeneity

Given that closure is satisfactorily achieved, the next most important consideration is twofold: (1) achieving a large enough number of captures to obtain reliable results, and (2) achieving a study for which the best model is the simplest possible one (e.g., Model M_0 rather than M_{tbh}, or Model M_b rather than M_{tb}). In this section, we discuss methods of eliminating variation of capture probabilities due to time, behavioral response to first capture, and heterogeneity of individuals.

Of the 3 factors that affect capture probabilities, time is the one most easily controlled by the biologist. He can select the season of the year the studies are to be conducted, the length of the trapping period, and the time of day when trapping is to be done. In all those decisions, the objective is to reduce variation in capture probabilities over time. Among other things, this means that equal effort should be expended on each occasion. For example, use the same number of traps thoughout, trap at the same time of day, and if bait is used, use the same type and amount on all occasions. The study should be done when weather conditions will be as constant as possible, because variable weather is likely to cause time variation in capture probabilities (Getz 1961).

Behavioral response is common in small mammals, and it is doubtful if much can be done to avoid it. Apparently for this reason, many biologists have studied and used strict removal methods for small mammal population estimation. If there is any choice, a method of capture (for livetrapping) should be used that will not result in a trap response. One approach is to use different methods of capture on each occasion (cf. Overton and Davis 1969). This is an excellent procedure, but its application to multiple capture occasions is severly limited because one cannot usually find 5 or 10 quite different capture methods. We note, however, that the use of different trapping methods probably will result in time variation (the different methods occur of necessity at different times). Consequently, there is a trade-off here: reduced behavioral variation may result in increased time variation.

A common source of heterogeneity is the lack of equal access to traps if traps are far apart relative to home range (Eberhardt 1969a). The above phenomena are part of the reasoning behind our recommendation of 4 traps per home range when we discuss sample size considerations in the next section.

Other sources of heterogeneity are differences in activity or catchability related to measurable characteristics such as size, age, or sex. Such sources can be removed by stratification if the attributes are recorded and sample size permits. Unfortunately, sample sizes are seldom adequate to allow stratification. Heterogeneity may also be due to some unrecognized attribute, and thus cannot be eliminated by stratification. Different methods of capture may increase the number of animals captured in this case, but will not eliminate the basic problem of heterogeneity of capture probabilities.

One possible method of eliminating heterogeneity, and possible trap response, is to locate traps randomly on each trapping occasion. We are unsure of the usefulness of the technique, and would like to see further studies conducted to see if significant reductions in heterogeneity result. For logistical reasons, the randomization of trap locations on each occasion probably is not feasible.

A valuable check on the livetrapping methods presented in this monograph can be made by using a second method of estimating the proportion of marked animals in the population. A common example is killtrapping or hunting after the capture–recapture experiment has been completed. Then, using the number of marked animals in the population, M_{t+1}, and the ratio of marked to unmarked animals obtained from the killtrapping or hunting, a second and somewhat independent population estimate is made with Chapman's (1951) version of the Lincoln estimate,

$$\hat{N} = \frac{(M_{t+1} + 1) \cdot (\text{all animals killtrapped} + 1)}{(\text{marked animals killtrapped} + 1)} - 1$$

(cf. Seber 1973). Of course traps are traps, and kill traps still present a strange object in the animal's environment. Another possibility for small mammals marked by toe clipping is to use tracks to obtain a ratio of marked to unmarked animals. Biologists have used smoked plates (Justice 1961) or dust covered tiles to obtain an estimate of the proportion of marked animals, and thus a second population estimate to compare with that obtained by livetrapping methods. The use of tracks has the distinct advantage of being a truly different method of sampling the population. Much confidence in the population estimate is obtained when the 2 estimates are similar.

In addition to design considerations, poststratification of the data can be used to create subsets of data which are more homogeneous with respect to capture probabilities. This is nothing more complex than partitioning the data into subsets on variables such as species, sex, age, weight, etc. If there are sufficient data, such poststratification is a valuable device. The subsets so created are then analyzed separately by the methods described in this monograph. The only additional testing one might do is for homogeneity among these strata (subsets) (see White 1975).

Sample Size

To obtain reliable estimates of population size, a sufficiently large sample must be taken. Typical sample size considerations are not applicable here (e.g., determining the numbers of plots to sample). Rather, "sample size" relates to the number of animals captured. For a live-trapping study, one must have both a large enough number of distinct animals captured and a sufficient number of recaptures (except for Model M_b and M_{bh}). The factors that control expected numbers of captures are (1) grid size, in terms of area covered and numbers of traps used, (2) capture probabilities, and (3) number of trapping occasions. We discuss these 3 factors in relation to the size of the experiment necessary to achieve precise population estimates.

The size of the grid is the first decision to be made. Grid size is a function of trap spacing, s, and the number of rows, r, and columns, c, of traps. Equal trap spacing and a systematic grid layout are suggested. At each grid station there will be one or more traps. If densities are very high, we recommend 2 or more traps per station to avoid competition for traps. Few studies have been made comparing 2 or more traps per station with 1 trap station, and we suggest further research is required to see where multiple traps per station increases the probability of capture for individual animals.

The objective of grid trapping over a short time period is to estimate the population size and usually also the density at the grid site. Because of "edge effect" (as discussed in the section on density estimation), it is necessary to estimate effective trapping area as well as N. In practice, this means we must be able to estimate strip width W as well as N. This requires that the traps be laid out to cover an area, rather than in a single line. Moreover, we must be able to associate each trap with an arbitrary X–Y coordinate system. For practical reasons, this implies some sort of regular grid layout (often square or rectangular) with equal spacing between traps. This latter aspect of trap layout is not necessary just for estimation of N: for example, if trapping was on a small island, or in an enclosure, knowledge of trap location in a coordinate system would not be needed. Finally, for the density estimate to be meaningful, the grid should be placed in a homogeneous habitat type, to assure uniform density over the grid.

For a choice of r and c when the objective is density estimation, we suggest that both values be greater than or equal to 5; as a minimum we recommend $r + c \geq 20$. Examples are a square grid 10×10 or a rectangle 5×15. We note much work in the literature relies on 16×16 grids (Gentry et al. 1968, Smith et al. 1971), and we suggest that grids should be at least that large for attempts to use the density estimation method based on nested subgrids. We base this recommendation on the fact that a large number of captures is required in each subgrid; hence, the larger the size of the subgrids the better the chance that large numbers of captures will be achieved.

The next decision to be made is the spacing of traps (i.e., the value of s). Most work with small mammals uses 15-m trap spacing or less (Barbehenn 1974, Smith et al. 1975). The rationale, when one is given, relates to the size of home range. Let s be the spacing between traps, and let 2W be the average linear home range size. Home ranges may rarely be circular, but assuming for design purposes that they may be circular, then their radius is W. We suggest at least 4 traps per home range. This implies $s \leq (\sqrt{2})W$. For best results we suggest $s \leq W/2$. Clearly, this implies some knowledge of home range sizes before a good study can be designed. This is not unreasonable to ask;

the biologist should have some behavioral knowledge of the species being studied so that sampling decisions can be made intelligently. In fact, in any statistical sampling problem, a good study cannot be planned without some prior knowledge of the population parameters to be estimated.

The analyses presented in this monograph for estimating population size require sufficient numbers of captures to produce satisfactory results. Defining "sufficient numbers" is an extremely complicated task. Based on our experience with both real and simulated data, however, some crude guidelines may be stated. For instance, experiments in which M_{t+1} is on the order of 10 or 20 animals simply do not provide enough information for the procedures discussed here to perform well. The number of different animals captured needs to be several times larger, and will depend heavily on the probabilities of capture of the population members being studied. That is, a population in which members have an "average" capture probability of 0.40 or 0.50 might only have to be as large as 50 before the estimation and testing techniques become useful, whereas a population size of 200 or so might require an average capture probability of only 0.20. For most studies, a relatively large number of recaptures must be realized before the experiment has a chance to produce useful results, and this again relates to the magnitude of the probabilities of capture involved. In general, the probabilities must be larger for smaller populations, but in no instance should N be less than 25 or average capture probabilities less than 0.10 when trapping small mammals for only a few occasions (say $t \leq 10$). These recommendations do not guarantee that the data can be satisfactorily analyzed, but we have seen enough real and simulated data to say that if the data fail these criteria it is improbable that a precise estimate will be achieved.

Estimation of density by the nested subgrid approach requires even larger sample sizes; the data on the smaller subgrids will otherwise be too sparse for reliable results. We believe that reliable density estimates using the subgrid approach require a grid of at least a 10 × 10 and as a *minimum* 75 to 100 different animals caught.

We now consider an approach to determining a minimum grid area on the basis of the above criteria. We start with the relationship $N = D \cdot A(W)$, where $A(W)$ is the effective trapping area. For a rectangular grid, this is

$$N = D[L_r L_c + 2(L_r + L_c)W + \pi W^2],$$

where L_r is the length of a row $[L_r = s \cdot (r-1)]$ and L_c $[L_c = s(c-1)]$ is the length of a column of traps. Thus, the area covered by the grid is $L_r L_c$.

As mentioned above, without some knowledge of D and W, a suitable study cannot be designed. Assume D_0 and W_0 are the best guesses of the values of the parameters. To determine if a grid study is *at all* feasible, set N = 50, substitute D_0 and W_0 in the above equation, set $L_r = L_c = L$ (a square grid), and solve for L:

$$L = \sqrt{0.8584(W_0)^2 + N/D_0} - 2W_0.$$

Then the actual area the grid must cover is L^2.

This procedure is not difficult; however, one must be careful to use the same basic units for D, L, and W. For example, let W_0 be 100 feet (30.5 m) and assume a density of 1 animal per acre (1/0.4 ha). For compatibility of units put D_0 in terms of square feet, then $D_0 = (1/43{,}560)$ feet^2. Solving for L gives

$$L = \sqrt{(0.8584)(100)^2 + (50)(43{,}560)} - 200$$
$$= 1{,}279 \text{ feet.}$$

This translates back into 37.5 acres (15.2 ha) as an absolute minimum grid size $[37.5 = (1{,}279)^2/43{,}560]$.

This is clearly conservative because not all animals will be caught. Improved planning requires us to determine grid size so that a given number of animals M_{t+1} will be caught. But the expected number of distinct animals caught depends upon the true underlying capture probabilities, which are not known. The only practical approach is to make the best guess at the overall average probability of first capture, $\bar{p}$, applicable during the study and then use the formula

$$M_{t+1} = N[1 - (1 - \bar{p})^t].$$

Because the value of t (number of occasions) has been introduced, the relevant computations for several realistic values of t can now be performed.

For example, assume $\bar{p} = 0.30$, set $M_{t+1} = 50$, and solve for N for several values of t. Given these values of N, solve for the value of L, and hence minimum grid size:

t	N	L	grid size in acres (ha)
4	66	1,498	51.5 (20.8)
6	57	1,378	43.6 (17.6)
8	53	1,323	40.0 (16.2)

In practice, this example means if trapping were only for 4 days, one would need a 16 × 16 grid, traps spaced 100 feet (30.5 m) apart. For an 8-day period of trapping, the same (expected) data could be obtained with a 14 × 14 grid of traps spaced 100 feet (30.5 m) apart.

This process can be reversed. Let us say a study is planned with a square grid of 16 traps, spaced 15 m apart (W = 7.5 m). Then $L_r = L_c = 225$ m. Substitution in the basic equation gives

$$N = D \cdot 57{,}552 \quad (m^2)$$

or

$$N = D \cdot 5.76 \quad (ha).$$

Assume further the study is to last 5 days and the average capture probability is about $\bar{p} = 0.30$. Then we have

$$M_6 = N(0.83).$$

We need to get at least 100 animals because the smallest subgrid requires 50, so the density should be large enough that $N \geqslant 120$, or

$$D \geqslant \frac{120}{5.76} = 20.8 \text{ animals/ha.}$$

Thus, to get reliable results in such a study we would say the true population density should equal or exceed 21 animals/ha. If the biologist has good reason to believe true density is only 10 or 12 animals/ha, the study is not even worth doing.

In addition to controlling the sampling effort through the size of the grid and the number of traps, the biologist can also select the number of trapping occasions. In theory, the more trapping times there are the better, but this ignores the fact that the closure assumption becomes less realistic as more time passes. We recommend a minimum of 5 trapping occasions, but 7 to 10 is better. The interval between occasions should be short. In practice, most trapping of small animals is either once a day (morning) or twice a day (morning and evening). Trapping only once a day is far less likely to introduce time variation. With morning and evening trapping there is very likely to be a difference in capture probabilities between times. If variation of behavior and heterogeneity should also be present, the correct model ends up as M_{tbh} for which no suitable estimator is available. Morning and evening trapping may, however, be aimed at different species. Then a workable design would be 5 (or 7) days of trapping in both morning and evening, but with separate analysis of the morning and evening data.

In removal studies, an absolute minimum is t = 3 occasions (*not* 2, as is often done) because it is impossible to test for equal capture probabilities when t = 2. We recommend that at least 4 removals be done.

Another valuable method of testing the adequacy of the design before going to the field is to simulate the experiment on a computer. Approximate parameter values can be chosen, and the experiment replicated as many times as necessary. Among the criteria that can be observed are selection of the appropriate model, the bias of selected estimators, and the achieved confidence level. Obviously, the validity of the simulations to the field

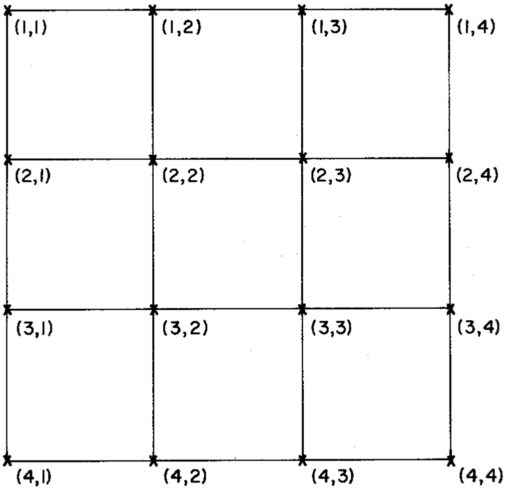


FIG. 20. Example of trap numbering for a 4 × 4 grid with a standard coordinate system.

study will depend on the similarity of parameters selected to the actual parameter values. However, insights about the experiment can be gained through simulations that cannot be achieved in any other way. We note that program CAPTURE has the capability to conduct these simulation studies.

Recording Data

If density is to be estimated on the basis of grid trapping data, the minimum information that must be taken when a capture occurs includes: (1) animal identification code, (2) trap location, and (3) trapping occasion. Usually, the species, sex, and age are also recorded for each animal, though for analyses given here it has no purpose except to partition the data by species, sex, and age. While analyzing the data separately by species is recommended, there often are not enough data to further partition by sex and age of animal.

For true removal studies (e.g., electrofishing), there is no animal identification code. As suggested by Raleigh (pers. comm.), it is very important in removal studies to record the individual by species and to analyze the data by at least major taxonomic groups.

In recapture studies, each animal should be marked uniquely, otherwise substantial information may be lost and it will be impossible to compute some of the tests for sources of variation. Though it should be obvious, we do mention that it is crucial to correctly record all data (e.g., animal number and trap number).

To know the trap location, traps must be numbered uniquely. Moreover, the biologist must know the relationship between the trap number and its coordinate on some (arbitrary) rectangular X–Y coordinate system. From the standpoint of data analysis, the best approach is to identify the traps in the field by these X–Y coordinates. We strongly recommend use of this system and we stress that density estimation using program CAPTURE requires data to be collected in the context of a coordinate system. We recommend that a corner trap be numbered (1, 1) then the rows become the "X axis" and columns the "Y axis." For example, a 4 × 4 grid would look like Fig. 20. This system can be extended to cover any regular rectangular grid of r rows and c columns.

When traps are checked both morning and evening, it is necessary to record not only the day of capture, but also the time.

Data recording will be facilitated by using standard field forms and standard conventions for trap numbering and animal identification. One example of a standardized method is presented in Brotzman and Giles (1966).

Data Anomalies

Various anomalies and unplanned events may occur in trapping. For example, (1) several animals may be found in 1 trap, (2) animals may be found dead in traps, (3) released animals may be found further down the grid trapped again on the same occasion, and (4) a trapped animal may escape when one attempts to remove it from the trap. We make the following suggestions regarding these happenings: (1) more than 1 animal per trap presents no problem. Record each animal separately. This type of data does not invalidate the analyses present here; (2) an animal dead in the trap in a livetrapping study is a more serious problem. If it is the last trapping occasion it does not matter. Otherwise the data analysis must be modified. Some of the methods described here can allow for known removals in a true livetrapping study (specifically Models M_0 and M_t). Because not all tests and estimators can be so modified, we have not dealt here with such modified models. We recommend the following: if trap deaths are less than 5 percent of total captures, remove those data from the total results, run the analyses and add that number of dead animals to $\hat{N}$, and then multiply the density estimate $\hat{D}$, by (1 + proportion dead). If such deaths are more than 20 percent of total captures, use the generalized removal method of analysis on first captures. For 5–20 percent trap deaths, the only safe analysis may be the removal method. These modifications give $\hat{N}$ and $\hat{D}$ relative to the first day of the study. If one desires these estimators to apply to the population remaining alive after the study, then simply delete all trap deaths from the data set prior to analysis; (3) the same animal is caught more than once on a given occasion, the only added information provided is on movement. We recommend that both captures be recorded, but only the results of the first capture be used for data analysis; and (4) an animal escapes during handling before it is tagged, or before the mark is read. Do nothing but record the fact. Do not attempt any sort of analysis of this "record." In essence, it does not become part of the data.

```
OVERALL TEST RESULTS --
Z-VALUE                                          -1.084
PROBABILITY OF A SMALLER VALUE                   .13925


TEST OF CLOSURE BY FREQUENCY OF CAPTURE.
(FREQUENCIES LESS THAN 10 ARE NOT COMPUTED.)

NUMBER OF CAPTURES  Z-VALUE  PROBABILITY
----------------------------------------

         2              .184      .57306
         3              .464      .67863
         4             1.208      .88650
```

FIG. 21a. Example of test procedure for population closure with the Scheme B taxicab data from Carothers (1973b).

COMPREHENSIVE EXAMPLES

Preceding sections have given the details of the specific models, estimators, and tests that are the basis for our analysis of capture data. The analysis of a set of livetrapping data by these methods will involve the model selection procedure, followed by estimation of N under the selected (or most appropriate model); density estimation may also be desired. Finally, in the course of these analyses various summary statistics can be computed. Below, we given several complete examples of the entire analysis process of livetrapping data for purposes of estimating population size. We do not include density estimation for all these examples because its essential features are always the same. Also, no further examples are given here of the analysis of removal data because that subject is substantially simpler than the analysis of capture–recapture data.

A Taxicab Example

Carothers (1973b) conducted an ingenious capture–recapture experiment on the taxicab population of Edinburgh, Scotland. Such a study has the advantages of known population size, yet the population is a real one (though not involving animals) as opposed to a computer simulation experiment. Two different sampling methods were used; we have already used Carothers' Scheme A in the section on Model M_h to illustrate the jackknife estimator (see Fig. 6); the entire 10 days of observations ("trappings") from Scheme B are used as an example here. In that scheme, observers had fixed stations in the city. This corresponds to a trapping study with 10 days of trapping at fixed trap locations. The true population size was 420, and we can

```
OCCASION            J=     1    2    3    4    5    6    7    8    9   10
ANIMALS CAUGHT   N(J)=    48   52   47   44   48   45   48   43   47   53
TOTAL CAUGHT     M(J)=     0   48   90  122  146  173  188  203  213  225  241
NEWLY CAUGHT     U(J)=    48   42   32   24   27   15   15   10   12   16
FREQUENCIES      F(J)=   104   67   51   12    6    1    0    0    0    0


 1. TEST FOR HETEROGENEITY OF TRAPPING PROBABILITIES IN POPULATION.
    NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(H)

           CHI-SQUARE VALUE =    7.913   DEGREES OF FREEDOM =  4   PROBABILITY OF LARGER VALUE =  .09482


 2. TEST FOR BEHAVIORAL RESPONSE AFTER INITIAL CAPTURE.
    NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(B)

           CHI-SQUARE VALUE =     .095   DEGREES OF FREEDOM =  1   PROBABILITY OF LARGER VALUE =  .75743


 3. TEST FOR TIME SPECIFIC VARIATION IN TRAPPING PROBABILITIES.
    NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(T)

           CHI-SQUARE VALUE =    2.247   DEGREES OF FREEDOM =  9   PROBABILITY OF LARGER VALUE =  .98693

 4. GOODNESS OF FIT TEST OF MODEL  M(H)
    NULL HYPOTHESIS OF MODEL  M(H) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(H)

           CHI-SQUARE VALUE =    2.300   DEGREES OF FREEDOM =  9   PROBABILITY OF LARGER VALUE =  .98578

           TEST OF MODEL M(H) BY FREQUENCY OF CAPTURE
           (FREQUENCIES LESS THAN 2T ARE NOT CALCULATED.)

           NUMBER OF CAPTURES  CHI-SQUARE  D.F.  PROBABILITY
           ----------------------------------------------------

                   1              6.769     9      .66113
                   2              6.246     9      .71504
                   3              4.714     9      .85847
5. GOODNESS OF FIT TEST OF MODEL  M(B)
   NULL HYPOTHESIS OF MODEL  M(B) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(B)

          CHI-SQUARE VALUE =    8.957   DEGREES OF FREEDOM = 16   PROBABILITY OF LARGER VALUE =  .91516

     5A. CONTRIBUTION OF TEST OF HOMOGENEITY OF FIRST CAPTURE PROBABILITY ACROSS TIME

          CHI-SQUARE VALUE =    7.598   DEGREES OF FREEDOM =  8   PROBABILITY OF LARGER VALUE =  .47369

     5B. CONTRIBUTION OF TEST OF HOMOGENEITY OF RECAPTURE PROBABILITIES ACROSS TIME

          CHI-SQUARE VALUE =    1.359   DEGREES OF FREEDOM =  8   PROBABILITY OF LARGER VALUE =  .99480


6. GOODNESS OF FIT TEST OF MODEL  M(T)
   NULL HYPOTHESIS OF MODEL  M(T) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(T)

          CHI-SQUARE VALUE =  188.341   DEGREES OF FREEDOM = 168   PROBABILITY OF LARGER VALUE =  .13483


7. TEST FOR BEHAVIORAL RESPONSE IN PRESENCE OF HETEROGENEITY.
   NULL HYPOTHESIS OF MODEL  M(H)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(BH)

          CHI-SQUARE VALUE =   15.148   DEGREES OF FREEDOM = 26   PROBABILITY OF LARGER VALUE =  .95458

MODEL SELECTION CRITERIA.  MODEL SELECTED HAS MAXIMUM VALUE.

MODEL      M(O)      M(H)      M(B)      M(BH)      M(T)      M(TH)      M(TB)      M(TBH)
CRITERIA    .93      1.00       .37       .52       0.00       .46        .36        .59
```

FIG. 21b. Example of model selection procedure based on Scheme B taxicab data from Carothers (1973b). Appropriate model probably is M_h. Suggested estimator is jackknife.

```
NUMBER OF TRAPPING OCCASIONS WAS              10
NUMBER OF ANIMALS CAPTURED, M(T+1), WAS      241
TOTAL NUMBER OF CAPTURES, N., WAS            475

FREQUENCIES OF CAPTURE,F(I)
   I=   1   2   3   4   5  6  7  8  9 10
F(I)= 104  67  51  12   6  1  0  0  0  0

          COMPUTED JACKKNIFE COEFFICIENTS

          N(1)      N(2)      N(3)      N(4)      N(5)
     1   1.900     2.700     3.400     4.000     4.500
     2   1.000      .289     -.878    -2.278    -3.722
     3   1.000     1.000     1.476     2.535     4.042
     4   1.000     1.000     1.000      .743      .077
     5   1.000     1.000     1.000     1.000     1.103


          THE RESULTS OF THE JACKKNIFE COMPUTATIONS

  I       N(I)    SE(I)    .95 CONF. LIMITS      TEST OF N(I+1) VS. N(I)
  0      241                                        CHI-SQUARE(1 D.F.)
  1      334.6    13.34    308.5     360.7           13.225
  2      370.2    21.53    328.0     412.4            2.344
  3      389.1    31.54    327.3     450.9            1.680
  4      408.6    44.09    322.1     495.0            2.137
  5      433.3    58.67    318.3     548.3            0.000

AVERAGE P-HAT =   .1301


INTERPOLATED POPULATION ESTIMATE IS        365 WITH STANDARD ERROR    20.3074

APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL        325 TO        406



HISTOGRAM OF F(I)

FREQUENCY     104   67   51   12    6    1    0    0    0    0
---------------------------------------------------------------
EACH * EQUALS    11 POINTS

       99       *
       88       *
       77       *
       66       *    *
       55       *    *    *
       44       *    *    *
       33       *    *    *
       22       *    *    *
       11       *    *    *    *    *
---------------------------------------------------------------
```

FIG. 21c. Example of population estimation with variable probability of capture by animal under Model M_h with Scheme B taxicab data from Carothers (1973b).

be reasonably certain of the closure assumption.

In this example, the concept of density is not applicable, nor is it meaningful to identify "trap" locations. Consequently, the first page of computer output is the test of closure (Fig. 21a). From Fig. 21a, we have the overall test value $z = -1.084$. This is not significant, consequently we would not reject closure. There were enough resightings 2, 3, and 4 times of the same taxicab to allow specific tests of closure on just those cabs seen that many times. None of the 3 tests are significant.

The next section of output is the model selection procedure (Fig. 21b). After the headings, some summary statistics are presented (daily captures n_j, cumulative marked M_j, new animals u_j, and capture frequencies f_j). The rest of the model selection procedure output (Fig. 21b) is devoted to the 7 tests of assumptions followed by the computed selection criteria and a suggested appropriate model. The first 3 tests compare Model M_o to Models M_h, M_b, and M_t, respectively. From Fig. 21b results are summarized below for those 3 tests:

Test of Model	*Chi-square value*	*Significance level*
M_o versus M_h	7.913	0.09482
M_o versus M_b	0.095	0.75740
M_o versus M_t	2.247	0.98693

The only indication of variability in capture probabilities is heterogeneity (significant at the 10% level). The study was designed to achieve constant numbers captured every day; therefore it is not surprising there is no indication of time variability in capture probabilities. Also, we do not find it surprising that there is no indication, from test 2, of behavioral response. One would not expect "trap" responses from taxicabs.

Tests 4, 5, and 6 examine the goodness of fit of Models M_h, M_b, and M_t, respectively. When any of those tests reject, we are saying that model does not appear appropriate for the given study. Test 4 indicates no departure from Model M_h (overall chi-square of 2.3 with 9 df). Nor does test 5 reject Model M_b (overall chi-square of 8.957 with 16 df). The goodness of fit test to Model M_t does not cause us to reject M_t, but does make us suspicious of that model (probability of a larger value = 0.13483). Finally, test 7 gives us no cause to suspect Model M_h should be rejected in favor of Model M_{bh} (chi-square of 16.148 with 27 df).

By itself, none of these 7 tests is definitive in suggesting the appropriate model; rather it is necessary to consider the results of all the tests. In this example we see some evidence of heterogeneity, but no evidence of time or behavioral variations in capture probabilities. Bearing in mind the robustness of the jackknife estimator of Model M_h, we are willing to conclude the appropriate model for these data is M_h. The "model selection criteria" computed by the program suggests M_h as appropriate (with Model M_o as the best second choice). Given these results, the user should look next at the jackknife estimate for these data and should consider all other estimators as being inappropriate.

Fig. 21c gives the results of estimation of N from Carothers' Scheme B data. This page of output will always give the summary statistics M_{t+1}, n., and the capture frequencies. Then the jackknife coefficients are given for the number of trapping occasions used. Next, the results of computing the first 5 jackknife estimates are shown. For example, $\hat{N}_{h1} = 334.6$ with a standard error of 13.34, while $\hat{N}_{h5} = 433.3$, and $\widehat{SE}(\hat{N}_{h5}) = 58.67$. In that example, the selected estimate of N is 365 with standard error of 20.3. The approximate 95 percent confidence interval on N is 325 to 406. That confidence interval fails to cover the true N = 420, reflecting the previously discussed poor coverage of the confidence intervals associated with the jackknife estimator, but the relative error of the estimator is only 13.1 percent (which is consistent with our simulation results on $\hat{N}_h$). In terms of real capture–recapture studies, this amount of bias is, in our opinion, acceptable.

Finally, we note that jackknife estimator is better (i.e., nearer to N = 420) for Carothers' Scheme B data than the commonly used estimators as Petersen (1896) or Schnabel (1938).

As a further aid in judging the validity of the study, the estimated average capture probability is given. From Fig. 21c, AVERAGE P-HAT = 0.1301. Our studies have indicated that a value of less than 0.10 suggests the capture results may not be trusted to produce good results. A 0.13 average probability is not very high, but it is acceptable when the true population size is as high as 420.

A Penned Rabbit Study

Edwards and Eberhardt (1967) reported the results of a livetrapping study on a penned population of 135 wild cottontails *Sylvilagus floridanus*. To our knowledge there have been few other controlled studies like that done, which is unfortunate because it would be very valuable to have more data sets on real populations where N is known.

```
OVERALL TEST RESULTS --
Z-VALUE                                   -2.132
PROBABILITY OF A SMALLER VALUE            .01650


TEST OF CLOSURE BY FREQUENCY OF CAPTURE.
(FREQUENCIES LESS THAN 10 ARE NOT COMPUTED.)

NUMBER OF CAPTURES  Z-VALUE  PROBABILITY
--------------------------------------


        2             3.691       .99989
```

FIG. 22a. Example of test for population closure procedure with cottontail data from Edwards and Eberhardt (1967).

It that study, 135 wild cottontails were captured and placed in a 40-acre (16.2 ha) rabbit-proof enclosure. After allowing 4 days for the rabbits to adjust to their new surroundings, livetrapping was conducted for 18 consecutive nights. When program CAPTURE was used to analyze the resultant data, the results were disappointing.

Fig. 22a shows the results of the closure test applied to Edwards and Eberhardt's (1967) data. Because $z = -2.132$ ($P = 0.0165$) one would normally question whether closure was true. In that example, we attribute the result to a time variation in daily capture probabilities, specifically there were fewer captures toward the end of the 18 days than at the start. As mentioned before, this "closure test" can detect only certain types of time variations of individual capture probabilities. It cannot of itself "know" the cause of the variations. Therefore, all the evidence in the data or otherwise available must be used to reach final conclusions about closure, or about other questions such as the presence of behavioral response (which also gets confounded with closure).

A brief data summary and the tests of assumptions are given by the model selection procedure (see Fig. 22b). From tests 1, 2, and 3, we see that Model M_0 is

```
OCCASION          J=     1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18
ANIMALS CAUGHT  N(J)=    9   8   9  14   8   5  18  11   4   3  16   5   2   7   9   0   4  10
TOTAL CAUGHT    M(J)=    0   9  15  18  29  33  34  44  51  52  53  62  62  63  68  74  74  74  76
NEWLY CAUGHT    U(J)=    9   6   3  11   4   1  10   7   1   1   9   0   1   5   6   0   0   2
FREQUENCIES     F(J)=   43  16   8   6   0   2   1   0   0   0   0   0   0   0   0   0   0   0


1. TEST FOR HETEROGENEITY OF TRAPPING PROBABILITIES IN POPULATION.
   NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(H)

        CHI-SQUARE VALUE =   11.110   DEGREES OF FREEDOM =   3   PROBABILITY OF LARGER VALUE =   .01115


2. TEST FOR BEHAVIORAL RESPONSE AFTER INITIAL CAPTURE.
   NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(B)

        CHI-SQUARE VALUE =     .062   DEGREES OF FREEDOM =   1   PROBABILITY OF LARGER VALUE =   .80367


3. TEST FOR TIME SPECIFIC VARIATION IN TRAPPING PROBABILITIES.
   NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(T)

        CHI-SQUARE VALUE =   46.932   DEGREES OF FREEDOM =  17   PROBABILITY OF LARGER VALUE =   .00012

4. GOODNESS OF FIT TEST OF MODEL  M(H)
   NULL HYPOTHESIS OF MODEL  M(H) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(H)

        CHI-SQUARE VALUE =   55.502   DEGREES OF FREEDOM =  17   PROBABILITY OF LARGER VALUE =   .00001

        TEST OF MODEL M(H) BY FREQUENCY OF CAPTURE
        (FREQUENCIES LESS THAN 2T ARE NOT CALCULATED.)

        NUMBER OF CAPTURES  CHI-SQUARE  D.F.  PROBABILITY
        -------------------------------------------------

               1               52.023    17       .00002

5. GOODNESS OF FIT TEST OF MODEL  M(B)
   NULL HYPOTHESIS OF MODEL  M(B) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(B)

        CHI-SQUARE VALUE =  102.913   DEGREES OF FREEDOM =  31   PROBABILITY OF LARGER VALUE =  0.00000

     5A. CONTRIBUTION OF TEST OF HOMOGENEITY OF FIRST CAPTURE PROBABILITY ACROSS TIME

        CHI-SQUARE VALUE =   47.065   DEGREES OF FREEDOM =  15   PROBABILITY OF LARGER VALUE =   .00004

     5B. CONTRIBUTION OF TEST OF HOMOGENEITY OF RECAPTURE PROBABILITIES ACROSS TIME

        CHI-SQUARE VALUE =   55.848   DEGREES OF FREEDOM =  16   PROBABILITY OF LARGER VALUE =   .00000

6. GOODNESS OF FIT TEST OF MODEL  M(T)
   NULL HYPOTHESIS OF MODEL  M(T) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(T)

        EXPECTED VALUES TOO SMALL.  TEST NOT PERFORMED.


7. TEST FOR BEHAVIORAL RESPONSE IN PRESENCE OF HETEROGENEITY.
   NULL HYPOTHESIS OF MODEL  M(H)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(BH)

        CHI-SQUARE VALUE =   52.023   DEGREES OF FREEDOM =  17   PROBABILITY OF LARGER VALUE =   .00002

MODEL SELECTION CRITERIA.  MODEL SELECTED HAS MAXIMUM VALUE.

MODEL       M(O)      M(H)      M(B)      M(BH)      M(T)      M(TH)      M(TB)      M(TBH)
CRITERIA     .15      0.00       .29       .01       1.00       .64        .22        .32
```

FIG. 22b. Example of model selection procedure based on cottontail data from Edwards and Eberhardt (1967). Appropriate model probably is M_t. Suggested estimator is Darroch.

not acceptable; moreover, the tests show that there is clearly time variation in average daily capture probabilities and that there probably is some heterogeneity (tests 3 and 1). Although test 2 fails to suggest that Model M_b is better than Model M_0, we cannot, from that alone, conclude there are no behavioral re-

```
OCCASION            J=    1    2    3    4    5    6    7    8    9   10   11   12   13   14   15   16   17   18
ANIMALS CAUGHT   N(J)=    9    8    9   14    8    5   18   11    4    3   16    5    2    7    9    0    4   10

TOTAL ANIMALS CAPTURED    76

          P-HAT(J)=   .12  .11  .12  .18  .11  .07  .24  .14  .05  .04  .21  .07  .03  .09  .12 0.00  .05  .13


POPULATION ESTIMATE IS        76 WITH STANDARD ERROR      .0165

APPROXIMATE 95 PERCENT CONFIDENCE INTERVAL        75 TO         77


HISTOGRAM OF N(J)

FREQUENCY       9    8    9   14    8    5   18   11    4    3   16    5    2    7    9    0    4   10
-----------------------------------------------------------------------------------------------------------
EACH * EQUALS   2 POINTS

      18                                    *
      16                                    *                   *
      14                   *                *                   *
      12                   *                *    *              *
      10        *         *    *                *    *              *                   *              *
       8        *    *    *    *    *           *    *              *              *    *              *
       6        *    *    *    *    *    *      *    *              *    *         *    *              *
       4        *    *    *    *    *    *      *    *    *    *    *    *         *    *         *    *
       2        *    *    *    *    *    *      *    *    *    *    *    *    *    *    *         *    *
-----------------------------------------------------------------------------------------------------------
```

FIG. 22c. Example of population estimation with time specific changes in probability of capture under Model M_t with cottontail data from Edwards and Eberhardt (1967).

sponses. In fact, test 7 ($p = 0.00002$) shows that there are time variations in capture probabilities which *may* be behavioral.

Only 2 of the goodness of fit tests (4 and 5) could be computed. The goodness of fit to Model M_t could not be done because of the small numbers of individuals caught most days. The other 2 tests, however, resulted in rejection, indicating that neither Model M_h nor M_b is a satisfactory model for the data.

When the goodness of fit test cannot be computed for Model M_t, the model selection criteria are computed based on the *assumption* that test 6 would give a significance level of 0.50. This tends to give undue weight to Model M_t, but there is really no good solution to this problem. For Edwards and Eberhardt's data, M_t is the indicated best fitting model, with Model M_{th} the next best. This is consistent with the observation that capture probabilities do appear affected by both time and heterogeneity.

The estimation of N based on Model M_t is given in Fig. 22c. The point estimate of population is 76, which is the same as the number of distinct animals captured (i.e., $M_{19} = 76$). As further evidence for determining whether $\hat{N}_t$ is a reliable estimate (and model), one should compute the overall average capture probability:

$$\bar{p} = \frac{n.}{t\hat{N}}.$$

For Edwards and Eberhardt's data:

$$\bar{p} = \frac{142}{(18)(76)} = 0.10380.$$

The advantage of using this formula instead of

$$\bar{p} = \left(\sum_{j=1}^{t} p_j\right)/(t)$$

is that we know the true population size is greater than or equal to 76. Thus, we know this value of p is an *upper bound* on the true expected value of p, and hence average capture probability is less than or equal to 0.10. This should raise a red flag to the investigator, average capture probability may well be less than 0.10,

```
ANIMAL  NUM.  MAXIMUM  AVERAGE  STANDARD   ANIMAL  NUM.  MAXIMUM  AVERAGE  STANDARD   ANIMAL  NUM.  MAXIMUM  AVERAGE  STANDARD
 I.D.   CAP.   DIST.    DIST.    ERROR      I.D.   CAP.   DIST.    DIST.    ERROR      I.D.   CAP.   DIST.    DIST.    ERROR
--------------------------------------    --------------------------------------    --------------------------------------

  268     6     4.5      1.7      .69        269     3     1.4       .7      .71        272     6     2.0       .8      .37
  273     6     1.4       .3      .28        274     5     1.0       .3      .25        276     6     1.4       .7      .29
  277     6     2.2      1.5      .30        279     5     8.5      4.0     1.75        280     4     2.2       .8      .42
  281     3     5.0      3.2     1.79        282     5     1.4      1.0     0.00        285     5     7.1      3.3     1.45
  286     6     2.0       .7      .29        287     5     2.0       .5      .29        288     6     2.2       .8      .20
  294     5     1.4       .4      .35        299     4     3.0      1.0     1.00        300     3     2.0      2.0     0.00
  163     4     1.0       .7      .33        165     4     4.1      3.5      .63        166     2     0.0      0.0     0.00
  167     4     2.0       .7      .67        169     4     2.2      1.2      .65        170     5     2.0       .5      .50
  171     5     4.1      1.8      .61        172     2     1.4      1.4     0.00        173     3     5.0      4.1      .92
  175     1     0.0      0.0     0.00        176     3     0.0      0.0     0.00        177     3     1.0      1.0     0.00
  184     1     0.0      0.0     0.00        187     4     2.2       .7      .75        188     3     0.0      0.0     0.00
  189     2     3.0      3.0     0.00        191     4     1.0       .3      .33        192     3     0.0      0.0     0.00
  193     3     0.0      0.0     0.00        196     3     1.4      1.2      .21        198     1     0.0      0.0     0.00
  199     2     5.0      5.0     0.00        200     2     2.2      2.2     0.00         86     2     0.0      0.0     0.00
   89     1     0.0      0.0     0.00         90     1     0.0      0.0     0.00         91     1     0.0      0.0     0.00
   92     2     1.0      1.0     0.00         93     2     0.0      0.0     0.00         95     2     4.1      4.1     0.00
   98     1     0.0      0.0     0.00        100     1     0.0      0.0     0.00        360     1     0.0      0.0     0.00
------------------------------------------------------------------------------------------------------------------------------
NOTE THAT AVERAGE DISTANCE ONLY REFERS TO DISTANCE BETWEEN SUCCESSIVE CAPTURES,
WHILE MAXIMUM DISTANCE REFERS TO THE GREATEST DISTANCE BETWEEN ANY TWO CAPTURE POINTS.
ALSO DISTANCE IS IN UNITS OF TRAP INTERVALS, I.E., IF THE INTER-TRAP DISTANCE IS 5 METERS,
AND THE MAX. DISTANCE IS 1.4, THEN THE MAX. DISTANCE IN METERS IS 1.4*5 OR 7 METERS.


SUMMARY BY FREQUENCY OF CAPTURE OF MAXIMUM DISTANCE BETWEEN CAPTURE POINTS.

                    NUMBER     SAMPLE    MEAN OF    STANDARD
                    CAPTURES    SIZE    MAX DIST.    ERROR
                    --------------------------------------

                       2          9       1.86       .621
                       3         10       1.58       .614
                       4          8       2.23       .360
                       5          8       3.45      1.018
                       6          7       2.25       .392
                    TOTAL        42       2.23       .575
                    --------------------------------------
```

FIG. 23a. Example of animal by animal summary of deer mouse capture data from V. Reid (pers. comm.).

in which case this analysis is not trustworthy when (apparent) population size is around 100.

In fact, if we use the true value of N = 135, then average capture probability in this experiment was 0.056. We also point out that even after 18 days of trapping only 53 percent of the population had been captured. This example illustrates our contention that it is very important to have average capture probabilities well above 0.05 or 0.10 for the population sizes typically encountered in capture–recapture studies (50 to 150). Considering that N = 135 and the population was penned, we suggest that the "true" situation was as follows. There was significant time variation and heterogeneity but little real behavioral response. The average daily capture probabilities declined in the last 9 (of the 18) days and this caused rejection of both the closure test and test 7. No good estimator of population size is available for data of this type and quality.

An Example of Trap Response

Many data sets we have seen on *Peromyscus maniculatus* have fit Model M_b (trap response). The following is an example supplied by V. Reid (pers. comm.). The data were taken in a 6-day livetrapping study near Wet Swizer Creek, Rio Blanco County, Colorado, August 1975. A rectangular grid of 9 × 11 traps was used, with 50-foot (15.2-m) trap spacing. One Sherman live trap (for small mammals) was placed at each grid point and trapping was done twice daily (morning and night); we have used only morning captures for this example. The reader should study Figs. 23a–e, which give the basic results, before proceeding.

A summary of movement information from recaptures is given in Fig. 23a. We remind the reader these distances are in units of 1 trap spacing [i.e., 50 feet (15.2 m) here]. Thus, the mean maximum distance of 2.23 implies the average maximum movement was 111.5 feet (34 m).

```
OCCASION          J=     1    2    3    4    5    6
ANIMALS CAUGHT N(J)=    16   29   27   29   32   38
TOTAL CAUGHT   M(J)=     0   16   31   38   44   48   51
NEWLY CAUGHT   U(J)=    16   15    7    6    4    3
FREQUENCIES    F(J)=     9    9   10    8    8    7


1. TEST FOR HETEROGENEITY OF TRAPPING PROBABILITIES IN POPULATION.
   NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(H)

       CHI-SQUARE VALUE =   37.489   DEGREES OF FREEDOM =  4   PROBABILITY OF LARGER VALUE =   .00000


2. TEST FOR BEHAVIORAL RESPONSE AFTER INITIAL CAPTURE.
   NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(B)

       CHI-SQUARE VALUE =   24.342   DEGREES OF FREEDOM =  1   PROBABILITY OF LARGER VALUE =   .00000


3. TEST FOR TIME SPECIFIC VARIATION IN TRAPPING PROBABILITIES.
   NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(T)

       CHI-SQUARE VALUE =   24.773   DEGREES OF FREEDOM =  5   PROBABILITY OF LARGER VALUE =   .00015


4. GOODNESS OF FIT TEST OF MODEL  M(H)
   NULL HYPOTHESIS OF MODEL  M(H) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(H)

       CHI-SQUARE VALUE =   25.225   DEGREES OF FREEDOM =  5   PROBABILITY OF LARGER VALUE =   .00013


5. GOODNESS OF FIT TEST OF MODEL  M(B)
   NULL HYPOTHESIS OF MODEL  M(B) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(B)

       CHI-SQUARE VALUE =    5.978   DEGREES OF FREEDOM =  8   PROBABILITY OF LARGER VALUE =   .64968

    5A. CONTRIBUTION OF TEST OF HOMOGENEITY OF FIRST CAPTURE PROBABILITY ACROSS TIME

       CHI-SQUARE VALUE =    1.135   DEGREES OF FREEDOM =  4   PROBABILITY OF LARGER VALUE =   .88866

    5B. CONTRIBUTION OF TEST OF HOMOGENEITY OF RECAPTURE PROBABILITIES ACROSS TIME

       CHI-SQUARE VALUE =    4.843   DEGREES OF FREEDOM =  4   PROBABILITY OF LARGER VALUE =   .30379


6. GOODNESS OF FIT TEST OF MODEL  M(T)
   NULL HYPOTHESIS OF MODEL  M(T) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(T)

       EXPECTED VALUES TOO SMALL.  TEST NOT PERFORMED.


7. TEST FOR BEHAVIORAL RESPONSE IN PRESENCE OF HETEROGENEITY.
   NULL HYPOTHESIS OF MODEL  M(H)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(BH)

       CHI-SQUARE VALUE =   22.217   DEGREES OF FREEDOM =  7   PROBABILITY OF LARGER VALUE =   .00233


MODEL SELECTION CRITERIA.  MODEL SELECTED HAS MAXIMUM VALUE.

MODEL        M(O)       M(H)       M(B)       M(BH)      M(T)       M(TH)      M(TB)      M(TBH)
CRITERIA      .26        .17       1.00        .65       0.00        .28        .55        .27
```

FIG. 23b. Example of model selection procedure based on deer mouse data from V. Reid (pers. comm.). Appropriate model probably is M_b. Suggested estimator is Zippin.

One estimate of W (strip width) is produced by taking half this average maximum distance (56 feet, 17 m, here), or half of it plus one intertrap distance (giving 106 feet, 32.3 m, here). Either approach probably would be an improvement over using W = 0, but these are essentially ad hoc approaches.

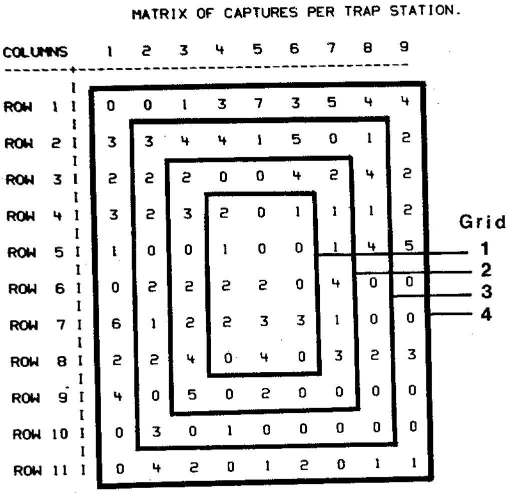
```
OCCASION            J=      1     2     3     4     5     6
TOTAL CAUGHT    M(J)=      0    16    31    38    44    48   51
NEWLY CAUGHT    U(J)=     16    15     7     6     4     3

ESTIMATED PROBABILITY OF CAPTURE, P-HAT =      .319119

ESTIMATED PROBABILITY OF RECAPTURE, C-HAT =  .677966


POPULATION ESTIMATE IS           56 WITH STANDARD ERROR      4.2859

APPROXIMATE 95 PERCENT CONFIDENCE INTERVALS            47 TO           65


HISTOGRAM OF U(J)


FREQUENCY       16    15     7     6     4     3
--------------------------------------------------
EACH * EQUALS     2 POINTS

        16       *     *
        14       *     *
        12       *     *
        10       *     *
         8       *     *     *
         6       *     *     *     *
         4       *     *     *     *     *     *
         2       *     *     *     *     *     *
--------------------------------------------------
```

FIG. 23c. Example of population estimation with constant probability removal estimator under Model M_b with deer mouse data from V. Reid (pers. comm.).

In this example, the test for closure gave $z = -1.586$ ($P = 0.05633$). But because the closure test is known to be biased when there is trap response, in the final analysis we conclude Model M_b adequately fits the data and closure is acceptable. A second bit of evidence on closure is given by the "ring" test (see Fig. 23e). That test examines for any apparent attraction of animals by the grid; it is not significant in this example, indicating that there is no basis to think immigration occurred.

From the model selection results (Fig. 23b) we find that tests 1, 2, and 3 all reject. This typically happens when there is strong trap response. Tests 4, 5, and 6 examine the goodness of fit of the 3 Models (M_h, M_b, and M_t, respectively). We see that Model M_h does not fit, but Model M_b does fit the data adequately. Unfortunately, the goodness of fit of M_t is the most difficult test to compute and with the small numbers of animals caught it could not be reliably computed.

Test 7 examines for possible behavioral variations in capture probabilities allowing for any heterogeneity which may be present (Fig. 23b). That test strongly suggests some form of behavioral response is present.



```
                 MATRIX OF CAPTURES PER TRAP STATION.

COLUMNS     1    2    3    4    5    6    7    8    9
--------+-------------------------------------------
        I
ROW   1 I   0    0    1    3    7    3    5    4    4
        I
ROW   2 I   3    3    4    4    1    5    0    1    2
        I
ROW   3 I   2    2    2    0    0    4    2    4    2
        I
ROW   4 I   3    2    3    2    0    1    1    1    2
        I
ROW   5 I   1    0    0    1    0    0    1    4    5
        I
ROW   6 I   0    2    2    2    2    0    4    0    0
        I
ROW   7 I   6    1    2    2    3    3    1    0    0
        I
ROW   8 I   2    2    4    0    4    0    3    2    3
        I
ROW   9 I   4    0    5    0    2    0    0    0    0
        I
ROW  10 I   0    3    0    1    0    0    0    0    0
        I
ROW  11 I   0    4    2    0    1    2    0    1    1
```

FIG. 23d. Nested subgrids used in the density estimation procedure with deer mouse data from V. Reid (pers. comm.). In the matrix, trap coordinates are rounded to the nearest whole integer. In the following goodness of fit tests, trap coordinates that are not integers and nonrectangular trapping grids will cause spurious results.

The suggested model is M_b, and the corresponding estimator of population size is, essentially, the Zippin removal estimator. (We have used the maximum likelihood estimator of N under the Zippin removal model. Recall that this differs slightly from the estimator usually computed.) Fig. 23c shows the estimate of population size under Model M_b to be $N = 56$, with a standard error of 4.3. The approximate 95 percent confidence interval on $\hat{N}$ is 51 to 65. Note that the program does not truncate the computed lower limit of 47 back to 51 (the number actually seen), but this could be done in reporting the results. The probability of initial capture is $\hat{p} = 0.32$ and the probability of recapture is $\hat{c} = 0.68$. These are significantly different because Model M_0 was rejected in favor of Model M_b. This sort of dramatic increase in capture probability after initial capture is entirely consistent with the properties of the data, for example, the observed increase over time in the n_j's (daily numbers captured).

The area covered by the trapping grid was 4.59 acres (1.86 ha); the naive esti-

```
        CHI-SQUARE TEST OF UNIFORM DENSITY BY ROWS.


ROW                    1         2         3         4         5         6         7         8         9        10
OBSERVED              27        23        18        15        12        12        18        20        11         4
EXPECTED          15.545    15.545    15.545    15.545    15.545    15.545    15.545    15.545    15.545    15.545
CHI-SQUARE         8.440     3.575      .388      .019      .809      .809      .388     1.276     1.329     8.575

ROW                   11
OBSERVED              11
EXPECTED          15.545
CHI-SQUARE         1.329

 TOTAL CHI-SQUARE =       26.94  WITH  10 DEGREES OF FREEDOM.  PROBABILITY OF LARGER VALUE =  .0027


         CHI-SQUARE TEST OF UNIFORM DENSITY BY COLUMNS.


COLUMN                 1         2         3         4         5         6         7         8         9
OBSERVED              21        19        25        15        20        18        17        17        19
EXPECTED          19.000    19.000    19.000    19.000    19.000    19.000    19.000    19.000    19.000
CHI-SQUARE          .211     0.000     1.895      .842      .053      .053      .211      .211     0.000

 TOTAL CHI-SQUARE =        3.47  WITH   8 DEGREES OF FREEDOM.  PROBABILITY OF LARGER VALUE =  .9012


        CHI-SQUARE TEST OF UNIFORM DENSITY BY RINGS (OUTER RING IS NUMBER 1).


RING                   1         2         3         4
OBSERVED              73        42        36        20
EXPECTED          62.182    48.364    34.545    25.909
CHI-SQUARE         1.882      .837      .061     1.348

 TOTAL CHI-SQUARE =        4.13  WITH   3 DEGREES OF FREEDOM.  PROBABILITY OF LARGER VALUE =  .2479
```

FIG. 23e. Chi-square tests of uniform density with deer mouse data from V. Reid (pers. comm.).

mate of density is thus 12.2 mice/acre (30.7/ha). Because of edge effect, we know this is an overestimate. As discussed above, movement data from recaptures is one basis for estimating the strip width W to get an estimate of effective trapping area. A conceptually better approach is to estimate W based on trapping grids of different sizes. Figs. 23e–f give results relevant to this approach for Reid's Wet Swizer Creek data. Four nested subgrids were chosen (as shown in Fig. 23d) as the basis of this procedure; from the number of captures per trap station, there is no apparent evidence of a nonuniform density over the grid. Figure 23e gives some logical tests of this uniform density assumption. However, our experience with the tests is that they often reject when there is, in fact, no real density gradient (i.e., systematic changes in density over the grid). Thus, even though there may be some variation in density by "rows," it is minor. Also, the choice of subgrids as shown tends to minimize the problem.

For each subgrid, an estimate of the population at risk of capture on that subgrid was obtained using Model M_b. As explained in the density estimation section, we can then estimate W. The results are shown in Fig. 23f. The 4 naive densities are shown, with the necessary computed constants (based on grid sizes), and the initial covariance matrix of the naive densitites.

The value of $\hat{W}$ is 105 with a (typically large) standard error of 58. Nonetheless, this value of $\hat{W}$ is significantly greater than zero, so we accept as our density estimate $\hat{D} = 5.54$ animals/acre (13.7/ha), ($\widehat{SE}(\hat{D}) = 1.92$).

When this subgrid approach to estimation of D fails, one must use

$$\tilde{D} = \hat{N}/A(\tilde{W}),$$

$$A(\tilde{W}) = A_g[1 + a\tilde{W} + b\tilde{W}^2],$$

```
          STARTING VALUES FOR DENSITY ESTIMATION--
          NUMBER OF GRIDS                          4
          TRAP INTERVAL                        50.00
          UNITS CONVERSION                  43560.00
          INITIAL DENSITY ESTIMATE           30.6995
          INITIAL STRIP WIDTH ESTIMATE         .5228

     GRID  NAIVE DENSITY    PERIMETER/AREA     PI/AREA         STARTING COVARIANCE MATRIX
      I       Y(I)               A(I)          B(I)

      1    32.67000         .3000000E-01    .1571E-03     150.
      2    61.71000         .1666667E-01    .5236E-04     507.      .152E+05
      3    17.42400         .1166667E-01    .2618E-04     7.74      232.       14.1
      4    12.19680         .9000000E-02    .1571E-04     1.16      34.7       2.11       .871


RESULTS OF ITERATIONS
    FUNCTION EVALUATIONS REQUIRED                        87
    ESTIMATED SIGNIFICANT DIGITS OF PARAMETER VALUES      8



          FITTED MODEL COMPARED TO THE DATA
          GRID(I)        Y(I)         F(I)
              1        32.670       32.651
              2        61.710       18.460
              3        17.424       13.941
              4        12.197       11.745

MULTIPLE CORRELATION COFFICIENT IS  .33718

ESTIMATED DENSITY=          5.538    1.9241   = ITS STANDARD ERROR
ESTIMATED STRIP WIDTH=    105.225   57.9903   = ITS STANDARD ERROR
CORRELATION OF ESTIMATORS      -.9824

TEST OF ESTIMATED STRIP WIDTH GREATER THAN ZERO.
Z-VALUE =  1.8145  PROBABILITY OF LARGER VALUE =   .0348


             FINAL COVARIANCE MATRIX

           150.0
           918.6      .1521E+05
           18.80      311.4       14.10
           3.359      55.62       2.518      .8714
```

FIG. 23f. Example of joint estimation of density and boundary strip width with deer mouse data from V. Reid (pers. comm.).

where A_g is the grid area, $\tilde{W}$ derives from either movement data or an independent source, and a and b are constants which can be computed. In fact, the program prints these constants on the density estimation page. From Fig. 23f, we have (for the total grid), a = 0.009 and b = 0.00001571.

As an illustration only, if we use $\tilde{W}$ = 55.75 (half the average maximum distance of 111.5) we get

$$A(\tilde{W}) = 4.59 \times 1.5506 = 7.117,$$

and hence,

$$\tilde{D} = \frac{56}{7.117} = 7.87.$$

(For previously given reasons, however,

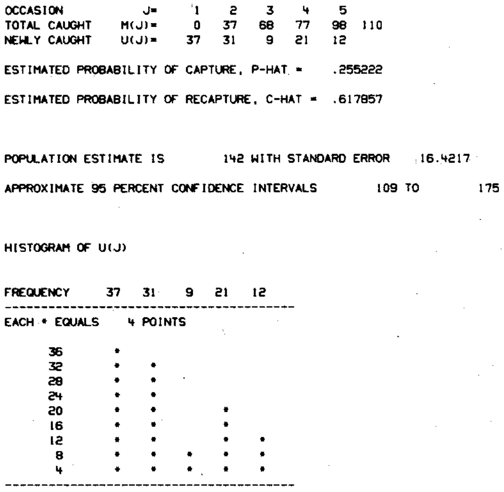
```
OCCASION            J=     1    2    3    4    5
ANIMALS CAUGHT   N(J)=    37   54   58   65   69
TOTAL CAUGHT     M(J)=     0   37   68   77   98  110
NEWLY CAUGHT     U(J)=    37   31    9   21   12
FREQUENCIES      F(J)=    34   20   28   15   13


 1. TEST FOR HETEROGENEITY OF TRAPPING PROBABILITIES IN POPULATION.
    NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(H)

           CHI-SQUARE VALUE =   49.016   DEGREES OF FREEDOM =   3   PROBABILITY OF LARGER VALUE =  0.00000


 2. TEST FOR BEHAVIORAL RESPONSE AFTER INITIAL CAPTURE.
    NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(B)

           CHI-SQUARE VALUE =   35.865   DEGREES OF FREEDOM =   1   PROBABILITY OF LARGER VALUE =  0.00000


 3. TEST FOR TIME SPECIFIC VARIATION IN TRAPPING PROBABILITIES.
    NULL HYPOTHESIS OF MODEL  M(O)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(T)

           CHI-SQUARE VALUE =   24.071   DEGREES OF FREEDOM =   4   PROBABILITY OF LARGER VALUE =   .00008

 4. GOODNESS OF FIT TEST OF MODEL  M(H)
    NULL HYPOTHESIS OF MODEL  M(H) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(H)

           CHI-SQUARE VALUE =   25.504   DEGREES OF FREEDOM =   4   PROBABILITY OF LARGER VALUE =   .00004

           TEST OF MODEL M(H) BY FREQUENCY OF CAPTURE
           (FREQUENCIES LESS THAN 2T ARE NOT CALCULATED.)

           NUMBER OF CAPTURES  CHI-SQUARE  D.F.  PROBABILITY
           ---------------------------------------------------

                   1               18.647     4      .00092
                   2               18.333     4      .00106
                   3                6.048     4      .19562
                   4                6.000     4      .19915

5. GOODNESS OF FIT TEST OF MODEL  M(B)
   NULL HYPOTHESIS OF MODEL  M(B) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(B)

          CHI-SQUARE VALUE =   13.510   DEGREES OF FREEDOM =   6   PROBABILITY OF LARGER VALUE =   .03562

    5A. CONTRIBUTION OF TEST OF HOMOGENEITY OF FIRST CAPTURE PROBABILITY ACROSS TIME

          CHI-SQUARE VALUE =    9.220   DEGREES OF FREEDOM =   3   PROBABILITY OF LARGER VALUE =   .02650

    5B. CONTRIBUTION OF TEST OF HOMOGENEITY OF RECAPTURE PROBABILITIES ACROSS TIME

          CHI-SQUARE VALUE =    4.289   DEGREES OF FREEDOM =   3   PROBABILITY OF LARGER VALUE =   .23187


6. GOODNESS OF FIT TEST OF MODEL  M(T)
   NULL HYPOTHESIS OF MODEL  M(T) VS. ALTERNATE HYPOTHESIS OF NOT MODEL  M(T)

          CHI-SQUARE VALUE =   84.121   DEGREES OF FREEDOM =  66   PROBABILITY OF LARGER VALUE =   .06562


7. TEST FOR BEHAVIORAL RESPONSE IN PRESENCE OF HETEROGENEITY.
   NULL HYPOTHESIS OF MODEL  M(H)  VS. ALTERNATE HYPOTHESIS OF MODEL  M(BH)

          CHI-SQUARE VALUE =   47.135   DEGREES OF FREEDOM =  10   PROBABILITY OF LARGER VALUE =   .00000

MODEL SELECTION CRITERIA.  MODEL SELECTED HAS MAXIMUM VALUE.

MODEL        M(O)       M(H)       M(B)       M(BH)      M(T)       M(TH)      M(TB)      M(TBH)
CRITERIA      .55        .42        .99        .89       0.00        .55        .88       1.00
```

we suspect that $\tilde{W} = 55.75$ is an underestimate and suggest $\hat{W} = 105$ is a better value to use.)

The sampling variance of $\tilde{D}$ is computable by the formula given in the density estimation section. The only missing term is $\widehat{\mathrm{Var}}(\tilde{W})$ which is obtained from the standard error of mean maximum distance (Fig. 23a)

$$\widehat{\mathrm{Var}}(\tilde{W}) = \left(\frac{50{*}0.575}{2}\right)^2 = 206.64.$$

Then using the formula for $\widehat{\mathrm{Var}}(\tilde{D})$ we get $\widehat{\mathrm{Var}}(\tilde{D}) = 0.98$, or $\widehat{\mathrm{SE}}(\tilde{D}) = 0.99$.



```
OCCASION          J=     1    2    3    4    5
TOTAL CAUGHT   M(J)=     0   37   68   77   98  110
NEWLY CAUGHT   U(J)=    37   31    9   21   12

ESTIMATED PROBABILITY OF CAPTURE, P-HAT =    .255222

ESTIMATED PROBABILITY OF RECAPTURE, C-HAT =  .617857


POPULATION ESTIMATE IS        142 WITH STANDARD ERROR    16.4217

APPROXIMATE 95 PERCENT CONFIDENCE INTERVALS        109 TO        175


HISTOGRAM OF U(J)


FREQUENCY     37   31    9   21   12
----------------------------------------
EACH * EQUALS    4 POINTS

       36     *
       32     *    *
       28     *    *
       24     *    *
       20     *    *         *
       16     *    *         *
       12     *    *         *    *
        8     *    *    *    *    *
        4     *    *    *    *    *
----------------------------------------
```

FIG. 24b. Example of population estimation with constant probability removal estimator under Model M_b with deer mouse data from S. Hoffman (pers. comm.).

An Example Where No Model Fits

In previous examples, the model selection procedure usually indicated a model that fit the data, and for which there was an estimator. However, with many real data sets we have found that no model fits the data. In those cases, the procedures given in this monograph serve to warn that no reliable estimator can be computed from the data (unless it can be judged the entire population has been caught).

In this example (Figs. 24a–e), we use livetrapping data collected by S. Hoffman (pers. comm.) in mid-July 1974 on deer mice. The study site was in a sagebrush–greasewood community at 4,500 feet (1,372 m) elevation in Curlew Valley, Idaho. A 12 × 12 grid of Sherman live traps, spaced 15 m apart, was used. Trapping was on 5 consecutive mornings, 110 individuals were caught (i.e., $M_{t+1} = 110$) and there were 283 captures in all.

Fig. 24a shows some summary statistics and the model selection procedure results. The daily captures increased (37, 54, 58, 65, 69), and although the new captures tended to decrease, they varied substantially (37, 31, 9, 21, 12). From these summary statistics, we would judge a substantial part of the catchable population remains untrapped. Phrased differently, there is no evidence here to suggest M_{t+1} is a reliable estimate of N.

From Fig. 24a, it can be seen that tests 1, 2, and 3 reject Model M_0 overwhelmingly. We thus conclude there is some type of variability in capture probabilities, and are suspicious that more than one source is operating. Test 4 also rejects the null hypothesis that Model M_h fits. At this point we can be fairly certain there is some form of time variability in daily capture probabilities, but it may be the often encountered result of behavioral variability.

The goodness of fit test to Model M_b indicates that this model is not a good fit to the data ($P = 0.03562$). Test 6 also suggests Model M_t is not an adequate fit to the data. Test 7 again confirms that some form of behavioral response is present so that Model M_h is inadequate. This leaves us uncertain as to how time and behavior variability are operating. Tests 5a and 5b

←

FIG. 24a. Example of model selection procedure based on deer mouse data from S. Hoffman (pers. comm.). Appropriate model probably is M_{tbh} or M_b. Suggested estimator is Zippin.

```
                    MATRIX OF CAPTURES PER TRAP STATION.

COLUMNS    1   2   3   4   5   6   7   8   9  10  11  12
-------+---------------------------------------------------
       I
ROW  1 I   4   3   5   5   5   3   2   0   3   4   3   1
       I
ROW  2 I   5   5   0   4   1   0   4   3   0   1   3   0
       I
ROW  3 I   4   2   0   3   1   0   2   2   1   3   3   1
       I
ROW  4 I   1   5   3   4   1   0   2   5   2   4   2   3
       I
ROW  5 I   0   1   4   3   0   1   3   0   0   1   5   2
       I
ROW  6 I   3   5   0   1   1   0   3   2   0   0   4   5
       I
ROW  7 I   2   0   2   0   0   0   1   3   0   1   2   1
       I
ROW  8 I   4   1   4   2   0   0   1   0   0   3   2   0
       I
ROW  9 I   1   2   4   2   4   0   0   2   4   4   2   4
       I
ROW 10 I   2   0   5   2   0   0   1   0   1   2   4   4
       I
ROW 11 I   1   3   1   0   0   0   0   3   4   2   2   2
       I
ROW 12 I   4   4   3   0   2   0   0   0   1   4   4   2


IN THE ABOVE MATRIX, TRAP COORDINATES ARE ROUNDED TO THE NEAREST WHOLE INTEGER.
IN THE FOLLOWING GOODNESS OF FIT TESTS, TRAP COORDINATES THAT ARE NOT INTEGERS
AND NON-RECTANGULAR TRAPPING GRIDS WILL CAUSE SPURIOUS RESULTS.

        CHI-SQUARE TEST OF UNIFORM DENSITY BY RINGS (OUTER RING IS NUMBER 1).

RING                 1          2          3          4          5          6
OBSERVED           107         76         53         32         11          4
EXPECTED        86.472     70.750     55.028     39.306     23.583      7.861
CHI-SQUARE       4.873       .390       .075      1.358      6.714      1.896

 TOTAL CHI-SQUARE =      15.31  WITH    5 DEGREES OF FREEDOM.  PROBABILITY OF LARGER VALUE =  .0091
```

FIG. 24c. Chi-square tests of uniform density with deer mouse data from S. Hoffman (pers. comm.).

help to clarify matters somewhat. Test 5a is the goodness of fit test to Zippin's (1956) constant probability removal model, i.e., the null hypothesis here is that first capture probability is constant. Test 5b examines the null hypothesis that recapture probabilities are constant. This latter test has $P = 0.23187$ suggesting that recapture probabilities may well be constant over time. By contrast with Test 5a ($P = 0.02650$) we conclude that first capture probabilities probably vary. These 7 tests suggest that behavior is the strongest factor affecting capture probabilities and that time is the next most significant factor.

The model selection criteria are a significant aid to interpreting data. In this example, the criteria values for models M_b, M_{bh}, M_{tb}, and M_{tbh} are 0.99, 0.89, 0.88, and 1.00, respectively. Recall that the most likely model is always given a value of 1, and the other criteria values are scaled accordingly, hence, we cannot say Model M_{tbh} actually fits the data. It is significant that these 4 models all account for behavioral response, thus corroborating our contention that there is a strong behavioral variation in capture probabilities. Because Model M_b is the next most likely model, one might select it as the basis for estimation. But there can be no

```
SUMMARY BY FREQUENCY OF CAPTURE OF MAXIMUM DISTANCE BETWEEN CAPTURE POINTS.

            NUMBER     SAMPLE   MEAN OF    STANDARD
            CAPTURES    SIZE    MAX DIST.   ERROR
            -------------------------------------

               2         20       1.46       .279
               3         28       2.43       .410
               4         15       2.26       .388
               5         13       2.44       .338
            TOTAL        76       2.14       .401
            -------------------------------------
```

FIG. 24d. Summary by frequency of capture of maximum distance between capture points.

strong reliance on $\hat{N}_b$ or any other estimator here, because none of the models fit the data. From Fig. 24b, the first capture probability (p) and the recapture probability under Model M_b are estimated to be

$$\hat{p} = 0.26 \text{ and } \hat{c} = 0.62.$$

These are known to be significantly different because test 2 rejected Model M_b. From Model M_b the estimate of N is $\hat{N}_b$ = 142, with an estimated standard error of 16.4. The approximate 95 percent confidence interval is 109–175. Other alternatives are also basically unsatisfactory. Model M_{bh} could be considered but it does not fit the data for any of its specific cases. This may be because first capture probabilities vary over time. Model M_h does not fit, and there are strong indications that behavior is the dominant source of variation. Based on these factors, we cannot recommend a valid estimation procedure for these data.

If density estimation is desired here, further problems arise. From Fig. 24c, there is no visual evidence of nonuniform density over the 12 × 12 grid. However, the ring test suggests significantly more animals were caught in the outer traps, which could be evidence of the grid "attracting" animals. This phenomenon does affect density estimation based on nested subgrids; in fact, this procedure failed for these data as the estimated W value was not significantly different from zero. The only recourse is to estimate W from movement data, or to use a value of W known to be representative for deer mice.

Fig. 24d shows the summary of maximum distances moved for recaptured animals. The overall average is 2.14 trap units. Converting this to meters (2.14 × 15) and dividing by 2 we have 16 m (52.7 feet) as a conservative estimate of W. This is typical of such values seen for deer mice on livetrapping grids with 15-m trap spacing. It is also known to be an underestimate. Using $\tilde{W}$ = 16 m would give a less biased density estimate than W = 0, but it would still probably result in an overestimate of D. The matter is further complicated if animals were in fact attracted to the grid. This phenomenon cannot be adequately dealt with, except by assessment lines (or designing a study so that no attraction occurs). Dropping the outer ring of traps and reanalyzing the data is another possibility, but then we are treating this outer ring as an assessment line.

```
OVERALL TEST RESULTS --
Z-VALUE                                  -2.142
PROBABILITY OF A SMALLER VALUE           .01610


TEST OF CLOSURE BY FREQUENCY OF CAPTURE.
(FREQUENCIES LESS THAN 10 ARE NOT COMPUTED.)

NUMBER OF CAPTURES  Z-VALUE  PROBABILITY
----------------------------------------

        2           -1.118      .13178
        3           -1.708      .04383
        4           -1.054      .14592
```

FIG. 24e. Example of test for population closure procedure with deer mouse data from S. Hoffman (pers. comm.).

Normally, we should have looked at the closure test early on in the model assessment procedure (see Fig. 24e); the results are z = −2.142, *P* = 0.0161. Because there is strong behavioral variation, we must discount this test; it cannot be relied upon. Thus, the only evidence we have of a closure failure is the ring test of Fig. 24c.

We propose the following as a plausible explanation of these data. From Hoffman (pers. comm.) we know there was no prebaiting of these traps, no time was allowed for the animals to become used to the traps. From the analysis we know there was behavioral response, and probably time variation in first capture probabilities (but not in recapture probabilities). The lack of prebaiting (often recommended in the literature) may have resulted in animals either increasing their first capture probabilities over time as the study progressed, or in an influx occurring during (rather than before) the study started. Prebaiting may solve the first problem. It will not solve the second problem of the grid attracting animals and thereby creating artificially high densities.

COMPREHENSIVE COMPUTER ALGORITHM

The computations necessary to calculate many of the estimates described in the preceding pages are quite lengthy, and in most cases, nearly impossible without a computer. Therefore, to provide methods of population estimation useful to the biologist, a comprehensive FORTRAN computer program, CAPTURE, has been written. The input to the program has been written in a free-form and natural style to make it easy to use.

The program is written in ANSI FORTRAN IV with several small exceptions so that it will function on most brands of digital computers. Statements known to cause compilation errors on IBM, CDC, Burroughs, Univac, Xerox, Honeywell, or DEC equipment have been flagged with comment statements, and the correct statement for the specific brand of computer included on a comment statement. A magnetic tape with the FORTRAN code and the 13 sets of example data illustrated in this monograph are available from

SHARE Program Library Agency
P. O. Box 12076
Research Triangle Park, N.C. 27709

at a cost of $40.00. Specifications for the tape (e.g., 7 or 9 track, 800 or 1,600 bpi, etc.) should be given when ordering the source program (No. 360D-17.5.002).

The program consists of a main routine and 54 subroutines. In addition, there are 7 common blocks. Included in the source code are ample comment statements to follow program flow. The total code consists of about 6,000 cards. Core requirements on an IBM 360 are about 200 K for the code without an overlay structure, but this can be reduced considerably with an overlay structure.

The dimensions of the program are presently set to allow up to 2,000 individual animals and 31 trapping occasions. The product of the number of captured animals and the number of trapping occasions must be less than 4,000. This will allow, for example, 30 trapping occasions and 80 animals, or 120 animals and 20 trapping occasions. These values can be changed by changing DIMENSION statements in the program.

SUMMARY

The results of this study provide methods for the estimation of animal population size N and density D from capture experiments. Both capture–recapture and removal studies to estimate population size are treated in detail. The primary focus of the work has been to relax the assumption of equal probability of capture. Three basic types of variation in probabilities of capture were examined: (1) capture probabilities vary with time, Model M_t, (2) capture probabilities vary by behavioral response, Model M_b, and (3) capture probabilities vary by individual animal, Model M_h. Models allowing these assumptions and various combinations of assumptions (i.e., Models M_{tb}, M_{th}, M_{bh}, and M_{tbh}) are treated. Population closure is assumed throughout.

An integrated approach was followed and the result is a blend of practical methods, examples of the analysis of real data, statistical theory, and results of computer simulation studies revealing some small sample properties of the methods. The estimation and testing problem has been treated in a standard and usually rigorous statistical framework. Above all, the necessity of assumptions, their importance, and statistical tests of their reasonableness have been emphasized. Assumptions must be carefully evaluated, both biologically and statistically, before a particular estimator can be used justifiably.

A sequence of 8 models that incorporate various assumptions has been developed. Point and interval estimators have been derived for most of the models. Furthermore, a comprehensive and objective (but not optimal) model selection strategy is provided. This is crucial because use of an incorrect model and improper assumptions is apt to produce biased estimates and incorrect inferences. In addition, proper interpretation of the test statistics and their interrelationships is shown to be somewhat difficult.

Although the practical analysis of data from capture experiments has been extended, additional research needs are apparent. Statistical testing within and between models will require more work. Models allowing other sets of assumptions need to be developed. Alternative estimation schemes (e.g., the generalized jackknife as opposed to the standard maximum likelihood approach) need attention, particularly for some of the more complex models. Additional work with incomplete contingency tables may prove fruitful (see Fienberg 1972). Interval estimates in general, and particularly interval estimates for the removal models, require additional research. Better tests for the important closure assumption are needed. Additional controlled studies are needed where population size is known, such as that reported by Edwards and Eberhardt (1967). Behavioral studies aimed at specific animal–trap interactions promise to provide interesting insights.

Research results indicate that acceptable estimates can be obtained if adequate data are collected properly. However, it is clear that biologists have not correctly conceptualized capture experiments over the past 4 decades. These limitations must be corrected if progress is expected in biological experiments involving capture of animals. Biologists have all too often gone into the field with only a few traps hoping to obtain meaningful data. Rarely will this be possible, even if only an estimate of population size is the goal. If density estimation is a goal, then further data requirements must be met.

As a reference point, it probably is appropriate to think in terms of 12 × 12, 16 × 16, or 20 × 20 square or rectangular grids and a trapping period of 8–9 days. Further guidelines on effort and sample size are given in the text. Our primary contention is that most capture experiments conducted in the past were quite inadequate in terms of design, effort, and sample size, which virtually prohibits justifiable inferences to be drawn from such studies.

Our computer simulation experiments have examined the small sample properties of the various point and interval estimators and the tests of model assumptions. The results have been informative and provide some basis for cautious optimism. Capture–recapture and removal methods have been overrated in the past and this probably has contributed to the lack of emphasis on design, sample size, and analytical methods. Estimation problems related to $\hat{D}$ and $\hat{N}$ represent difficult subjects. Our results provide hope that a rigorous analysis may often allow useful inferences to be drawn if future experiments are well designed and provide adequate capture and recapture or removal data for analysis.

LITERATURE CITED

ANDRZEJEWSKI, R., AND W. JEZIERSKI. 1966. Studies on the European hare: XI. Estimation of population density and attempt to plan the yearly take of hares. Acta Theriol. 11:433–448.

ANONYMOUS. 1971. Mathematical intrinsics: Information manual. Burroughs Corp. El Monte, Ca. 37 pp.

ARNASON, A. N. 1972a. Prediction methods and variance estimates for the parameters of the triple catch–two population model with migration and death. Univ. Manitoba Sci. Rept. 54:1–31.

———. 1972b. Parameter estimates from mark–recapture experiments on two populations subject to migration and death. Res. Pop. Ecol. 13(2):97–113.

———. 1973. The estimation of population size, migration rates and survival in a stratified population. Res. Pop. Ecol. 15(1):1–8.

———, AND L. BANIUK. 1977. User's manual: POPAN-2, A data maintenance and analysis system for recapture data. Univ. Manitoba, Dept. Computer Sci., Winnipeg, Canada.

BAILEY, G. N. A. 1968. Trap-shyness in a woodland population of bank voles *(Clethrionomys glareolus)*. J. Zool. 156(10):517–521.

BAILEY, N. T. J. 1951. On estimating the size of mobile populations from recapture data. Biometrika 38(3/4):293–306.

BARBEHENN, K. R. 1974. Estimating density and home range size with removal grids: The rodents and shrews of Guam. Acta Theriol. 19:191–234.

BIRNBAUM, A. 1969. Concepts of statistical evidence. *In* S. Morganbesser, et al. (Eds.). Philosophy, science and method: Essays in honor of E. Nagel. St. Martins Press, New York, N.Y.

BRAATEN, D. O. 1969. Robustness of the DeLury population estimator. J. Fish. Res. Bd. Can. 26(2):339–355.

BRENT, R. P. 1973. Algorithms for minimization without derivatives. Prentice-Hall, Inc., Englewood Cliffs, N.J. 196 pp.

BROTZMAN, R. L., AND R. H. GILES, JR. 1966. Electronic data processing of capture–recapture and related ecological data. J. Wildl. Manage. 30(2):286–292.

BROWNIE, C., D. R. ANDERSON, K. P. BURNHAM, AND D. S. ROBSON. 1978. Statistical inference from band recovery data: A handbook. U.S. Fish Wildl. Serv., Resour. Publ. 131. 212 pp.

BURNHAM, K. P., AND W. S. OVERTON. 1969. A simulation study of live-trapping and estimation of population size. Ore. St. Univ., Dept. Stat., Tech. Rept. No. 14. 60 pp. and appendix.

CAROTHERS, A. D. 1971. An examination and extension of Leslie's test of equal catchability. Biometrics 27(3):615–630.

———. 1973a. The effects of unequal catchability on Jolly-Seber estimates. Biometrics 29(1):79–100.

———. 1973b. Capture–recapture methods applied to a population with known parameters. J. Anim. Ecol. 42(1):125–146.

CHAPMAN, D. G. 1948. A mathematical study of confidence limits of salmon populations calculated from sample tag ratios. Int. Pac. Salmon Fish. Comm. Bull. II:69–85.

———. 1951. Some properties of the hypergeometric distribution with applications to zoological sample censuses. Univ. Cal. Publ. Stat. 1(7):131–160.

———. 1952. Inverse, multiple and sequential sample censuses. Biometrics 8(4):286–306.

———. 1954. The estimation of biological populations. Ann. Math. Stat. 25(1):1–15.

COOLEY, W. W., AND P. R. LOHNES. 1962. Multivariate procedures for the behavioral sciences. John Wiley & Sons, Inc., New York, N.Y. 211 pp.

CORMACK R. M. 1966. A test for equal catchability. Biometrics 22(2):330–342.

———. 1968. The statistics of capture–recapture methods. Oceanogr. Mar. Biol. Annu. Rev. 6:455–506.

CRAIG, C. C. 1953. On the utilization of marked specimens in estimating populations of flying insects. Biometrika 40(1/2):170–176.

CROWCROFT, P., AND J. N. R. JEFFERS. 1961. Variability in the behavior of wild house mice *(Mus musculus* L.) towards live-traps. Proc. Zool. Soc. Lond. 137(4):573–582.

DARROCH, J. N. 1958. The multiple-recapture census: I. Estimation of a closed population. Biometrika 45(3/4):343–359.

———. 1959. The multiple-recapture census: II. Estimation when there is immigration or death. Biometrika 46(3/4):336–351.

———. 1961. The two-sample capture–recapture census when tagging and sampling are stratified. Biometrika (3/4):241–260.

DAVIS, D. E. 1956. Manual for analysis of rodent populations. Edwards Brothers, Inc., Ann Arbor, Mich. 82 pp.

DELURY, D. B. 1947. On the estimation of biological populations. Biometrics 3(4):145–167.

———. 1951. On the planning of experiments for the estimation of fish populations. J. Fish. Res. Bd. Can. 8(4):281–307.

———. 1958. The estimation of population size by marking and recapture procedure. J. Fish. Res. Bd. Can. 15(1):19–25.

DICE, L. R. 1938. Some census methods for mammals. J. Wildl. Manage. 2(2):119–130.

———. 1941. Methods for estimating populations of animals. J. Wildl. Manage. 5(4):398–407.

EBERHARDT, L. L. 1969a. Population estimates from recapture frequencies. J. Wildl. Manage. 33(1):28–39.

———. 1969b. Population analysis. Pp. 457–495. *In* R. H. Giles, Jr., (Ed.). Wildlife management techniques. 3rd ed. The Wildlife Society, Wash., D.C.

———, T. J. PETERLE, AND R. SCHOFIELD. 1963. Problems in a rabbit population study. Wildl. Monogr. No. 10:1–51.

EDWARDS, W. R., AND L. L. EBERHARDT. 1967. Estimating cottontail abundance from live trapping data. J. Wildl. Manage. 31(1):87–96.

FELLER, W. 1950. An introduction to probability theory and its applications. Vol. 1. John Wiley & Sons, Inc., New York, N.Y. 509 pp.

FIENBERG, S. E. 1972. The multiple recapture census for closed populations and incomplete 2^k contingency tables. Biometrika 59(3):591–603.

FISCHLER, K. J. 1965. The use of catch-effort, catch sampling, and tagging data to estimate a population of blue crabs. Trans. Amer. Fish. Soc. 94(3):287–310.

FISHMAN, G. S. 1973. Concepts and methods in discrete event digital simulation. Wiley-Interscience, New York, N.Y. 385 pp.

FLETCHER, R. 1972. FORTRAN subroutines for minimization of quasi-Newton methods. Atom. Energy Res. Estab., Rep. 7125. Harwell, Eng. 34 pp.

FLYGER, V. R. 1959. Movements and home range of the gray squirrel *Sciurus carolinensis*, in two Maryland woodlots. Ecology 41(3):365–369.

GASKELL, T. J., AND B. J. GEORGE. 1972. A Bayesian modification of the Lincoln index. J. Anim. Ecol. 41(3):377–384.

GEIS, A. D. 1955a. Trap response of the cottontail rabbit and its effect on censusing. J. Wildl. Manage. 19(4):466–472.

GENTRY, J. B., F. B. GOLLEY, AND M. H. SMITH. 1968. An evaluation of the proposed International Biological Program census methods for estimating small mammal populations. Acta Theriol. 13:313–327.

GETZ, L. L. 1961. Response of small mammals to live-traps and weather conditions. Amer. Midl. Nat. 66(1):160–169.

GILBERT, R. O. 1973. Approximations of the bias in the Jolly-Seber capture–recapture model. Biometrics 29(3):501–526.

GOODMAN, L. A. 1953. Sequential sampling tagging for population size problems. Ann. Math. Stat. 24(1):56–69.

GRAY, H. L., AND W. R. SCHUCANY. 1972. The generalized jackknife statistic. Marcel Dekker, New York, N.Y. 308 pp.

GRAYBILL, F. A. 1976. Theory and application of the linear model. Duxbury Press, North Scituate, Mass. 704 pp.

GREEN, R. G., AND C. A. EVANS. 1940. Studies on a population cycle of snowshoe hares on the Lake Alexander area: I. Gross annual census, 1932–1939. J. Wildl. Manage. 4(2):220–238.

HANSSON, L. 1969. Home range, population structure and density estimates at removal catches with edge effect. Acta Theriol. 14:153–160.

———. 1974. Influence area of trap stations as a function of number of small mammals exposed per trap. Acta Theriol. 19:19–25.

HAYNAM, G. E., Z. GOVINDARAJULU, AND F. C. LEONE. 1970. Tables of the cumulative non-central chi-square distribution. Pp. 1–42. *In* H. L. Harter and D. B. Owen (Eds.). Selected tables in mathematical statistics. Vol. 1. Amer. Math. Soc., Providence, R. I.

HAYNE, D. W. 1949a. Two methods for estimating populations from trapping records. J. Mammal. 30(4):399–411.

———. 1949b. Calculation of size of home range. J. Mammal. 30(1):1–18.

HUBER, J. J. 1962. Trap response of confined cottontail populations. J. Wildl. Manage. 26(2):177–185.

HUDSON, D. J. 1971. Interval estimation from the likelihood function. J. R. Stat. Soc., B. 33:256–262.

IMSL 1976. International mathematical statistical libraries. 5th Ed. Houston, Texas. (pages not consecutively numbered).

JACKSON, C. H. N. 1933. On the true density of tsetse flies. J. Anim. Ecol. 2(2):204–209.

———. 1937. Some new methods in the study of *Glossina morsitans*. Proc. Zool. Soc. Lond., 1936. 811–896.

———. 1939. The analysis of an animal population. J. Anim. Ecol. 8(2):238–246.

———. 1940. The analysis of a tsetse fly population. Ann. Eugen. 10(4):332–369.

JENNRICH, R. I., AND F. B. TURNER. 1969. Measurement of non-circular home range. J. Theoret. Biol. 22:227–237.

JOLLY, G. M. 1965. Explicit estimates from capture–recapture data with both death and immigration—stochastic model. Biometrika 52(1/2):225–247.

JUSTICE, K. E. 1961. A new method for measuring home ranges of small mammals. J. Mammal. 42(4):462–470.

KAUFMAN, D. W., G. C. SMITH, R. M. JONES, J. B. GENTRY, AND M. H. SMITH. 1971. Use of assessment lines to estimate density of small animals. Acta Theriol. 16:127–147.

KENDALL, M. G., AND W. R. BUCKLAND. 1971. Dictionary of statistical terms. 3rd edition. Hafner Publ. Co., New York, N.Y. 166 pp.

———, AND A. STUART. 1973. The advance theory of statistics. Vol. 2, 3rd ed. Griffin, London, Eng. 723 pp.

KETCHEN, K. S. 1953. The use of catch-effort and tagging data in estimating a flat-fish population. J. Fish. Res. Bd. Can. 10(4):459–483.

LESLIE, P. H. 1952. The estimation of population parameters from data obtained by means of the capture-recapture method. II. The estimation of total numbers. Biometrika 39(3/4):363–388.

———. 1958. Statistical appendix. J. Anim. Ecol. 27(1):84–86.

———, AND D. H. S. DAVIS. 1939. An attempt to determine the absolute number of rats on a given area. J. Anim. Ecol. 8(1):94–113.

———, AND D. CHITTY. 1951. The estimation of population parameters from data obtained by means of the capture–recapture method: I. The maximum likelihood equations for estimating the death-rate. Biometrika 38(3/4):269–292.

———, ———, AND H. CHITTY. 1953. The estimation of population parameters from data obtained by means of the capture–recapture method: III. An example of the practical applications of the method. Biometrika 40(1/2):137–169.

LINCOLN, F. C. 1930. Calculating waterfowl abundance on the basis of banding returns. Circ. U.S. Dept. Agric. No. 118:1–4.

MACLULICH, D. A. 1951. A new technique of animal census, with examples. J. Mammal. 32(3):318–328.

MANLY, B. F. J. 1970. A simulation study of animal population estimation using the capture-recapture method. J. Appl. Ecol. 7(1):13–39.

———. 1971. Estimates of a marking effect with capture-recapture sampling. J. Appl. Ecol. 8(1):181–189.

MARTEN, G. G. 1970. A regression method for mark-recapture estimation of population size with unequal catchability. Ecology 51(2):291–295.

MILLER, R. P. 1964. Simultaneous statistical inference. McGraw-Hill Book Co., New York, N.Y. 272 pp.

MOHR, C. O., AND W. A. STUMPF. 1966. Comparison of methods for calculating areas of animal activity. J. Wildl. Manage. 30(2):293–304.

MOOD, A. M., F. A. GRAYBILL, AND D. C. BOES. 1974. Introduction to the theory of statistics. 3rd ed. McGraw-Hill Book Co., New York, N.Y. 564 pp.

MORAN, P. A. P. 1951. A mathematical theory of animal trapping. Biometrika 38(3/4):307–311.

NIE, N. H., C. H. HULL, J. G. JENKINS, K. STEINBRENNER, AND D. H. BRENT. 1975. Statistical package for the social sciences. 2nd ed. McGraw-Hill Book Co., New York, N.Y. 675 pp.

NIXON, C. N., W. R. EDWARDS, AND L. L. EBERHARDT. 1967. Estimating squirrel abundance from live trapping data. J. Wildl. Manage. 31(1):96–101.

O'FARRELL, M. J., D. W. KAUFMAN, AND D. W. LUNDAHL. 1977. Use of live-trapping with the assessment line method for density estimation. J. Mammal. 58(4):575–582.

OMAND, D. N. 1951. A study of populations of fish based on catch-effort statistics. J. Wildl. Manage. 15(1):88–98.

OVERTON, W. S., AND D. E. DAVIS. 1969. Estimating the number of animals in wildlife populations. Pp. 403–455. *In* R. H. Giles, Jr., (Ed.). Wildlife management techniques. 3rd ed. The Wildlife Society, Wash., D.C.

PAULIK, G. J. 1963. Estimates of mortality rates from tag recoveries. Biometrics 19(1):28–57.

———, AND D. S. ROBSON. 1969. Statistical calculations for change-in-ratio estimators of population parameters. J. Wildl. Manage. 33(1):1–27.

PETERSEN, C. G. J. 1896. The yearly immigration of young plaice into the Limfjord from the German Sea. Rep. Danish Biol. Sta. 6:1–48.

POLLOCK, K. H. 1975. A K-sample tag-recapture model allowing for unequal survival and catchability. Biometrika 62(3):577–583.

———, D. L. SOLOMON, AND D. S. ROBSON. 1974. Tests for mortality and recruitment in a K-sample tag-recapture experiment. Biometrics 30(1):77–87.

PUCEK, Z. 1969. Trap response and estimation of numbers of shrews in removal catches. Acta Theriol. 14:403–426.

QUENOUILLE, M. 1949. Approximate tests of correlation in time series. J. Roy. Stat. Soc., Ser. B. 11(1):68–84.

———. 1956. Notes on bias in estimation. Biometrika 43(3/4):353–360.

RICKER, W. E. 1958. Handbook of computations for biological statistics of fish populations. Bull. Fish. Res. Bd. Can. No. 119. 300 pp.

———. 1975. Computation and interpretation of biological statistics of fish populations. Dept. Environ., Bull 191. 382 pp.

ROBSON, D. S. 1969. Mark-recapture methods of population estimation. Pp. 120–140. *In* N. L. Johnson and H. Smith, Jr. (Eds.). New developments in survey sampling. Wiley-Interscience, New York. N.Y.

———, AND W. A. FLICK. 1965. A nonparametric statistical method for culling recruits from a mark–recapture experiment. Biometrics 21(4):936–947.

———, AND H. A. REGIER. 1964. Sample size in Petersen mark–recapture experiments. Trans. Amer. Fish. Soc. 93(3):215–226.

———, AND ———. 1968. Estimation of population number and mortality rates. Pp. 124–158. *In* W. E. Ricker (Ed.). Methods for assessment of fish production in fresh waters. IBP Handbook No. 3, Blackwell Scientific Publ., Oxford, Eng.

———, AND W. D. YOUNGS. 1971. Statistical analysis of reported tag-recaptures in the harvest from an exploited population. Biometrics Unit Rep. BU-369-M, Cornell Univ., Ithaca, N.Y. 15 pp.

SAMUEL, E. 1968. Sequential maximum likelihood estimates of the size of a population. Ann. Math. Stat. 39(3):1057–1068.

SCHNABEL, Z. E. 1938. The estimation of the total fish population of a lake. Amer. Math. Month. 45(6):348–352.

SCHUMACHER, F. X., AND R. W. ESCHMEYER. 1943. The estimate of fish population in lakes or ponds. J. Tenn. Acad. Sci. 18(3):228–249.

SEBER, G. A. F. 1962. The multi-sample single recapture census. Biometrika 49(3/4):339–349.

———. 1965. A note on the multiple-recapture census. Biometrika 52(1/2):249–259.

———. 1970. The effects of trap-response on tag-recapture estimates. Biometrics 26(1):13–22.

———. 1973. Estimation of animal abundance and related parameters. Griffin, London, Eng. 506 pp.

———, AND J. F. WHALE. 1970. The removal method for two and three samples. Biometrics 26(3):393–400.

SMITH, H. D., C. D. JORGENSEN, AND H. D. TOLLEY. 1972. Estimation of small mammals using recapture methods: Partitioning of estimator variables. Acta Theriol. 17:57–66.

SMITH, M. H. 1968. A comparison of different methods of capturing and estimating numbers of mice. J. Mammal. 49(4):455–462.

———, J. B. GENTRY, AND F. B. GOLLEY. 1969. A preliminary report on the examination of small mammal census methods. Pp. 25–29. *In* K. Petrusewicz and L. Ryszkowski (Eds.). Energy flow through small mammal populations. Polish Scientific Publ., Warsaw, Poland.

———, R. BLESSING, J. G. CHELTON, J. B. GENTRY, F. B. GOLLEY, AND J. T. MCGINNIS. 1971. Determining density for small mammal populations using a grid and assessment lines. Acta Theriol. 16:105–125.

———, R. H. GARDNER, J. B. GENTRY, D. W. KAUFMAN, AND M. H. O'FARREL. 1975. Density estimation of small mammal populations. Pp. 25–53. *In* F. B. Golley, K. Petrusewicz and L. Ryszkowski (Eds.). Small mammals: Their productivity and population dynamics. Intern. Biol. Prog. 5., Cambridge Univ. Press, London, Eng.

SNEDECOR, G. W., AND W. G. COCHRAN. 1967. Statistical methods. 6th ed. Iowa St. Univ. Press, Ames, Ia. 593 pp.

SOUTHWOOD, T. R. E. 1966. Ecological methods. Methuen Co., London, Eng. 391 pp.

STEIN, C. 1962. A remark on the likelihood principle. J. Roy. Stat. Soc., A. 125(4):565–568.

STICKEL, L. F. 1954. A comparison of certain methods of measuring ranges of small mammals. J. Mammal. 35(1):1–15.

SWIFT, D. M., AND R. K. STEINHORST. 1976. A technique for estimating small mammal population densities using a grid and assessment lines. Acta Theriol. 21(4):471–480.

SWINEBROAD, J. 1964. Net-shyness and wood thrush populations. Bird-Banding 33(3):196–202.

TABER, R. D., AND I. M. COWAN 1969. Capturing and marking wild animals. Pp. 277–317. *In* R. H. Giles, Jr. (Ed.). Wildlife management techniques. 3rd ed. The Wildlife Society, Wash., D.C.

TANAKA, R. 1951. Estimation of vole and mouse populations on Mount Ishizuchi and on the uplands of Southern Shikoku. J. Mammal. 32(4):450–458.

———. 1952. Theoretical justification of the mark-and-release index for small mammals. Bull. Kochi Women's Coll. 1:38–47.

———. 1956. On differential response to live traps of marked and unmarked small mammals. Ann. Zool. Jap. 29(1):44–51.

———. 1963. On the problem of trap-response types of small mammal populations. Res. Pop. Ecol. 2:139–146.

———. 1970. A field study of the effect of prebaiting on censusing by the capture-recapture method in a vole population. Res. Pop. Ecol. 12(1):111–125.

———. 1972. Investigation into the edge effect by use of capture-recapture data in a vole population. Res. Pop. Ecol. 13(2):127–151.

———, AND M. KANAMORI. 1967. New regression formula to estimate the whole population for recapture-addicted mammals. Res. Pop. Ecol. 9(1):83–94.

TANTON, M. T. 1965. Problems of live-trapping and population estimation for the wood mouse, *Apodemus sylvaticus* (L.). J. Anim. Ecol. 34(1):1–22.

WHITE, E. G. 1975. Identifying population units that comply with capture–recapture assumptions in an open community of alpine grasshoppers. Res. Pop. Ecol. 16(2):153–187.

YOUNG, H., J. NEESS, AND J. T. EMLEN, JR. 1952. Heterogeneity of trap response in a population of house mice. J. Wildl. Manage. 16(2):169–180.

ZIPPIN, C. 1956. An evaluation of the removal method of estimating animal populations. Biometrics 12(2):163–189.

———. 1958. The removal method of population estimation. J. Wildl. Manage. 22(1):82–90.