

MODEL SELECTION IN CAPTURE-RECAPTURE EXPERIMENTS

R. M. Cormack

University of St. Andrews  
and Cornell University

BU-712-M

August 1980

Abstract

This note describes the use of a GLIM program to select an appropriate, parsimonious model for data from a standard capture-recapture experiment.

## MODEL SELECTION IN CAPTURE-RECAPTURE EXPERIMENTS

Capture-recapture studies of animal populations can be analyzed by GLIM if the complete capture histories of all animals seen during the study are noted. If a sighting at the  $i^{\text{th}}$  sample is indicated by a value 1 for the  $i^{\text{th}}$  suffix, a non-sighting by the value 2, the data  $n_{abc\dots}$  are entered in the standard order for a  $2^S$  factorial design:  $n_{111\dots}, n_{211\dots}, n_{121\dots}, n_{221\dots}, \dots$ . The unknown number of unseen animals  $n_{222\dots}$  is given a fictitious value, whose effect is nullified by giving it zero weight  $W$ .

A closed population is represented by main effects only (Fienberg 1972). Trap dependence lasting for one period only is represented by the corresponding two-factor interaction. In the GLIM parameterization, death between one pair of samples generates one non-zero interaction, as also does birth between one pair of samples, a different pattern of interactions being generated by each different generalization of the model. Since these patterns are non-hierarchical, the explanatory variables cannot be entered as factors but vectors must be constructed.

The complete program for a 5-sample experiment with no loss on capture is as follows, the illustrative data being taken from Manly and Parr (1968).

```

$UNITS 32          $DATA N          $READ
1 0 5 1 0 1 2 5 0 4 2 7 4 2 1 19
1 1 0 3 0 2 0 9 9 4 1 13 10 13 21 100
$CAL A=%GL(2,1)-1 : B=%GL(2,2)-1 : C=%GL(2,4)-1 :
      D=%GL(2,8)-1 : E=%GL(2,16)-1 : W=1 :
AB=A*B : BC=B*C : CD=C*D : DE=D*E :
B3=AB*C : B4=B3*D : D2=DE*C : D1=D2*B :
PBD=B+C+D : PB=PBD+E : PD=PBD+A
$EDIT 32 W 0 $YVAR N $ERR P $WEI W
$FIT

```

A selection of models can be fitted in any sequence determined jointly by biological and statistical considerations:

A+B+C+D+E            for closed population

add AB+B<sub>3</sub>+B<sub>4</sub>        for birth

add DE+D<sub>1</sub>+D<sub>2</sub>        for death

add AB+BC+CD+DE      for trap dependence

Constant effort : in a closed population is GM alone ;

in a population with birth only replaces B+C+D+E by PB ;

in a population with death only replaces A+B+C+D by PD ;

and with birth and death replaces B+C+D by PBD .

Out-of-range birth or death rates are revealed by negative estimates for the corresponding parameters. Such parameters should be omitted from the fit.

In assessing the deviance from a model, care has to be taken if there are many zeroes in the data set, since every zero necessarily contributes 0 to the deviance. Subtracting the number of zeroes from the degrees of freedom is recommended, although this may overcompensate. The generalized Pearson  $\chi^2$ , obtained by `$PRINT %X2`, gives an alternative indication of fit. The difference between models seems to be reasonably assessed by the appropriate  $\chi^2$ .

If none of the sequences of models for birth, death and trap dependence are acceptable, then heterogeneous behaviour of individuals, or other serious breakdown of the basic capture-recapture assumptions, may be suspected. The pattern of standardized residuals may provide some clue to the nature of this breakdown.

For the illustrative data birth and death are both identified as occurring, while trap dependence between successive periods is unimportant. Constant effort is unacceptable as a model. The deviance from the standard birth and death Jolly-

Seber model is however 37.1 with 19 d.f. Examination of the residuals reveals  $n_{12111}$  to have a positive standardized residual of 4.6. Repetition of the analysis with this observation removed [EDIT 3 W 0] reveals  $n_{12211}$  as the only large residual, again positive. After its removal and the removal of the then negative birth parameter, the deviance is 18.2 with 18 d.f. We may suspect that temporary emigration affected the burnet moths after the first sampling occasion.

Some further details are to be found in Cormack (1981) and in another paper to be submitted shortly for publication. If one of the Jolly-Seber models are found to be appropriate, the usual estimates are readily obtained from the GLIM estimates.

R. M. Cormack  
University of St. Andrews  
and Cornell University

- Cormack, R. M. (1981). Loglinear models for capture-recapture experiments on open populations. In *The Mathematical Theory of the Dynamics of Biological Populations*, ed. R. W. Hiorns and D. Cooke. Academic Press.
- Fienberg, S. E. (1972). The multiple recapture census for closed populations and incomplete  $2^k$  contingency tables. *Biometrika*, 59, 591-603.
- Manly, B. F. J. and Parr, M. J. (1968). A new method of estimating population size, survivorship, and birth rate from capture-recapture data. *Trans. Soc. Brit. Ent.*, 18, 81-89.