Program LINLOGN

Users Instructions

James E. Hines[1], Kenneth H. Pollock[2] and James D. Nichols[1]

[1] U. S. Fish & Wildlife Service
Patuxent Wildlife Research Center
Laurel, MD. 20708

[2] Dept. of Statistics
N.C. State University
Box 8203
Raleigh, NC. 27695-8203

Institute of Statistics
Mimeograph Series No.
North Carolina State University
Raleigh, North Carolina
November, 1984

Introduction:

Pollock et al. (1984) described models incorporating auxiliary variables for use in capture-recapture and removal experiments. In some of these models the auxiliary variables (e.g. environmental variables or effort expended in obtaining the trapped sample) are associated with sampling occasion, and capture probability is modeled as a function of these variables using a linear-logistic relationship. Maximum likelihood estimation of population size under these models requires iterative techniques. Program LINLOGN was thus developed to compute population estimates and test statistics for 3 models.

The 3 models incorporated in LINLOGN are all removal models (Pollock et al. 1984:331), differing only in the way in which capture probability is modeled. Under model M2, the 2-parameter model, capture probability is a constant over all sampling periods and is expressed as:

$$P = \frac{e^{b_0}}{1 + e^{b_0}}$$

where P is the capture probability, and

b_0 is a model parameter.

Under M3, the 3-parameter model, capture probability for each sampling period is expressed as a function of an auxiliary variable:

$$P_i = \frac{e^{b_0 + b_1 Y1_i}}{1 + e^{b_0 + b_1 Y1_i}}$$

where P_i is the capture probability in the i-th time period,

b_0 and b_1 are model parameters, and

$Y1_i$ is the value of the auxiliary variable for the i-th time period.

Under M4, the 4-parameter model, capture probability for each sampling period is expressed as a function of 2 auxiliary variables:

$$P_i = \frac{e^{b_0 + b_1 Y1_i + b_2 Y2_i}}{1 + e^{b_0 + b_1 Y1_i + b_2 Y2_i}}$$

where P_i is the capture probability in the i-th time period,
b_0, b_1 and b_2 are model parameters,
$Y1_i$ is the value of the first auxiliary variable for the i-th time period, and
$Y2_i$ is the value of the second auxiliary variable for the i-th time period.

Program Input:

Program input consists of 6* records as described in the following table:

Variable	Type	Record Number	Columns	Description
TITLE	REAL*8	1	1-72	72 Character title of data set
K	INTEGER	2	1-2	Number of sampling periods
PRECIS	REAL*8	2	4-13	Convergence criterion for the parameter vector containing b_0, b_1, and b_2. Convergence is assumed when the distance between successive parameter vectors < PRECIS.
PRECSN	REAL*8	2	14-23	Convergence criterion for the parameter N. Convergence is assumed when the distance between successive N's < PRECSN.
MAXITR	INTEGER	2	25-30	Maximum number of iterations permitted.
IOPT	INTEGER	2	32	Option code which, when set to zero, causes the program to use $Y1_i^2$ in place of $Y2_i$.

Input variables (cont.):

Variable	Type	Record Number	Columns	Description
U(I)	REAL*8	3	10 cols each	Number of unmarked animals captured or removed in sample period I.
Y1(I)	REAL*8	4	10 cols each	First auxiliary variable for sample period I.
Y2(I)	REAL*8	5	10 cols each	Second auxiliary variable for sample period I.
N	REAL*8	6	1-8	Starting value for population size.

*More records are required if there are more than 10 time periods.

Other Variables:

Additional variables used in program LINLOGN are listed below:

Variable	Type	Description
THETA	REAL*8(4)	Parameter vector which contains the values of b_0, b_1, b_2, and N.
INFO	REAL*8(4,4)	Information matrix which contains second order partial derivatives of the log-likelihood function with respect to the parameters.
G	REAL*8(4)	Gradient vector which contains first order partial derivatives of the log-liklihood function with respect to each parameter.

Computing Algorithm:

Estimates are computed for each model using the method of scoring (see outline in Brownie et al. 1978:211-212). The basic procedure is as follows:

1) Set vector THETA to initial values.
2) Perform steps 3-7 until distance between successive THETA's < PRECIS.
3) Compute vector G.
4) Compute INFO.
5) Compute the inverse of INFO.
6) Compute new THETA using G, inverse of INFO, and old THETA.
7) Compute distance between old THETA and new THETA.
8) Compute log-likelihood function for final values of THETA.
9) Print THETA, INFO, and log-likelihood function.
10) Perform a "goodness-of-fit" test by comparing the observed values of U(I) versus the expected values.

Program LINLOGN was originally written using this method exactly, but in use we found the models were very slow to converge. We believe this was because N is a discrete parameter and differs in magnitude greatly from the other parameters. We solved this problem by taking N out of the parameter vector, THETA, and solving for N separately. The new method can be described as follows:

A) Set N to an initial value.
B) Perform steps C-D until the distance between the old N and the new N $\leq$ PRECSN.
C) Perform steps 1-7 (above), where THETA no longer includes N.
D) Compute new N using final values from step C.
E) Perform steps 8-10 (above).

Program Output:

Output from program LINLOGN consists of a page of definitions and input listing, and 1 or 2 pages of results for each of the three models. For each of the models, the program prints:

1) The formula for capture probability (P) specific to the model.
2) Initial values of the parameter vector (THETA).
3) Number of iterations before convergence.
4) Final values of the parameter vector.
5) Standard errors of each of the parameters.
6) Capture probabilities using final values of the parameters.
7) Variance-covariance matrix of the parameters.
8) Log-likelihood function evaluated for the final values of the parameters.
9) Likelihood ratio test results comparing the model to the previous model (Not applicable for the first model).
10) Goodness of fit test results.

Note that (5) and (7) are maximum likelihood estimates. Sometimes it may be appropriate to scale these as suggested by Pollock et al. (1984).

Example:

In this example we consider data, presented by Paloheimo (1963), from a Canadian lobster fishery at Port Maitland in 1950-1951. A listing of the input data appears in appendix 1, and the output appears in appendix 2. Note that the standard errors do not match those in Pollock et al. (1984). This was due to scaling of the standard errors.

REFERENCES

Brownie, C. , D. R. Anderson, K. P. Burnham, and D. S. Robson (1978). Statistical Inference from band recovery data-A handbook. *U.S. Dep. Inter., Fish and Wildl. Serv. Resour. Publ.* **131**. ,211-212.

Paloheimo, J. E. (1963). Estimation of catchabilities and population sizes of lobsters. *Journal of the Fisheries Research Board of Canada* **20**, 59-88.

Pollock, K. H., J. E. Hines and J. D. Nichols (1984). The Use of Auxiliary Variables in Capture-Recapture and Removal Experiments. *Biometrics* **40**, 329-340.

Appendix 1. Example input data from Paloheimo (1963)

```
** LOBSTER POP. AT PORT MAITLAND, CANADA, 50-51, PALOHEIMO(1963)
13     .1D-09      .01   32766
    604    495    282    207    119    156    132    254    299    325
    247    276    222
 33.664 27.743 17.254 14.764 11.190 16.263 14.757 32.922 45.519 43.523
 37.478 43.367 37.960
    7.9    7.7    6.3    3.5    3.1    2.9    3.1   3.25    3.4    3.6
    4.0    5.9    6.1
 8388.
```

Appendix 2. Example output using data in appendix 1.

```
PROGRAM LINLOGN (11/08/84)

     PROGRAM LINLOGN COMPUTES MAXIMUM LIKELIHOOD ESTIMATES OF POPULATION SIZE USING 3 NESTED CAPTURE-RECAPTURE/REMOVAL MODE
     DESCRIBED BY POLLOCK, HINES AND NICHOLS (1984; BIOMETRICS 40: 329-340).  THE MODELS INCORPORATE 2, 1, AND 0 AUXILIARY
     VARIABLES, RESPECTIVLY.

VARIABLE                 DEFINITION
--------     ----------------------------------------
   K      -  NUMBER OF SAMPLING PERIODS
  M(I)    -  NUMBER OF MARKED ANIMALS AVAILABLE IN SAMPLE I (NUMBER CAUGHT BEFORE I)
  U(I)    -  NUMBER OF UNMARKED ANIMALS CAUGHT IN SAMPLE I
  E(I)    -  EXPECTED VALUE OF U(I)
  Y1(I)   -  FIRST AUXILIARY VARIABLE IN SAMPLE I

  Y2(I)   -  SECOND AUXILIARY VARIABLE IN SAMPLE I
                         ** LOBSTER POP. AT PORT MAITLAND, CANADA, 50-51, PALOHEIMO(1963)

NUMBER OF SAMPLING PERIODS (K) =  13
PRECISION AT WHICH TO STOP ITERATIVE PROCEDURE (PRECIS,PRECIS(N)) =      .1000E-09      .1000E-01
MAXIMUM NUMBER OF ITERATIONS BEFORE ABORTING =  32766
```

DATA SUMMARY STATISTICS

M	U	Y1	Y2
.00000000	604.00000000	33.66400000	7.90000000
604.00000000	495.00000000	27.74300000	7.70000000
1099.00000000	282.00000000	17.25400000	6.30000000
1381.00000000	207.00000000	14.76400000	3.50000000
1588.00000000	119.00000000	11.19000000	3.10000000
1707.00000000	156.00000000	16.26300000	2.90000000
1863.00000000	132.00000000	14.75700000	3.10000000
1995.00000000	254.00000000	32.92200000	3.25000000
2249.00000000	299.00000000	45.51900000	3.40000000
2548.00000000	325.00000000	43.52300000	3.60000000
2873.00000000	247.00000000	37.47800000	4.00000000
3120.00000000	276.00000000	43.36700000	5.90000000
3396.00000000	222.00000000	37.96000000	6.10000000
3618.00000000			

Appendix 2. Example output using data in appendix 1. (Cont)

M4 - 4 PARAMETER MODEL
****

$$P(I) = \frac{EXP(B0 + B1*Y1(I) + B2*Y2(I))}{1 + EXP(B0 + B1*Y1(I) + B2*Y2(I))}$$

** LOBSTER POP. AT PORT MAITLAND, CANADA, 50-51, PALOHEIMO(1963)

INITIAL VALUES OF THETA (B0,B1,B2,N)

B0	=	.000000000000000
B1	=	.000000000000000
B2	=	.000000000000000
N	=	8388.000000000000010

FINAL VALUES OF THETA (STANDARD ERROR) AFTER 444 ITERATIONS

B0	=	-3.936974820617689	(.090568358489308)
B1	=	.030363532995430	(.002042767611226)
B2	=	.110365730489030	(.010043669999242)
N	=	5499.743155414821110	(238.900881461207483)

CAPTURE PROB. (P) =

.114768963 .095800406 .061932413 .043008055 .037145019 .042163908 .041218880 .070524652 .101596184 .098132753
.086467731 .122491858 .108018378

THETA VARIANCE-COVARIANCE MATRIX

.008202627559	.000000174677	-.000661999918	-14.173626393560
.000000174677	.000004172900	-.000006903266	-.301335680719
-.000661999918	-.000006903266	.000100875307	1.103523286149
-14.173626393560	-.301335680719	1.103523286149	57073.631162941907

****
** LOG LIKELIHOOD FUNCTION = 17106.558935396035900 **
****

GOODNESS OF FIT TEST FOR THE 4 PARAMETER MODEL

I	U(I)	E(I)	CHI-SQUARE
1	604.	631.20	1.17210
2	495.	466.41	1.75271
3	282.	272.63	.32170
4	207.	177.60	4.86626
5	119.	146.79	5.26229
6	156.	160.44	.12277
7	132.	150.23	2.21195
8	254.	246.44	.23167
9	299.	329.98	2.90922
10	325.	286.35	5.21609
11	247.	227.55	1.66188
12	276.	294.48	1.16009
13	222.	227.88	.15161

TOTAL CHI SQUARE = 27.04
DEGREES OF FREEDOM= 9.
OVERALL PROB. = .0014

Appendix 2. Example output using data in appendix 1. (Cont.)

M3 - 3 PARAMETER MODEL
****

```
P(I) =   EXP(B0 + B1*Y1(I))
       ----------------------
       1 + EXP(B0 + B1*Y1(I))
```

** LOBSTER POP. AT PORT MAITLAND, CANADA, 50-51, PALOHEIMO(1963)

INITIAL VALUES OF THETA

```
      B0 =            .000000000000000
      B1 =            .000000000000000
      N  =        5499.743155414821110
```

FINAL VALUES OF THETA (STANDARD ERROR) AFTER 149 ITERATIONS

```
      B0 =          -3.293655243609094 (            .0542020435105231)
      B1 =            .037019273554173 (            .0018264373538191)
      N  =        4722.697268283490160 (         109.1544839058129671)
```

CAPTURE PROB. (P) =

```
 .114311471  .093924656  .065686126  .060250488  .053180858  .063470202  .060235817  .111559785  .166786560  .156769285
 .129403080  .156007384  .131426595
```

THETA VARIANCE-COVARIANCE MATRIX

.002937861521	-.000059338024	-1.866234849489
-.000059338024	.000003335873	-.103191350313
-1.866234849489	-.103191350313	11914.701356744382

```
*************************************************************
** LOG LIKELIHOOD FUNCTION =        17045.117612402782500 **
*************************************************************

**************************************************
**          LIKELIHOOD RATIO TEST               **
**   4 PARAMETER MODEL VS. 3 PARAMETER MODEL    **
** CHI-SQUARE VALUE =   122.8826                **
** PROB. =      .0000 WITH 1 DEGREE OF FREEDOM  **
**************************************************
```

GOODNESS OF FIT TEST FOR THE 3 PARAMETER MODEL

I	U(I)	E(I)	CHI-SQUARE
1	604.	539.86	7.62075
2	495.	392.87	26.54856
3	282.	248.95	4.38812
4	207.	213.35	.18889
5	119.	176.97	18.98834
6	156.	199.98	9.67050
7	132.	177.74	11.77058
8	254.	309.35	9.90468
9	299.	410.90	30.47407
10	325.	321.81	.03171
11	247.	223.99	2.36435
12	276.	235.09	7.11767
13	222.	167.15	17.99557

```
TOTAL CHI SQUARE  =     147.06
DEGREES OF FREEDOM=      10.
OVERALL PROB.     =     .0000
```

M2 - 2 PARAMETER MODEL
****

$$P(I) = \frac{EXP(B0)}{1 + EXP(B0)}$$

** LOBSTER POP. AT PORT MAITLAND, CANADA, 50-51, PALOHEIMO(1963)

INITIAL VALUES OF THETA (B0,N)
B0 = .000000000000000
N = 4722.697868283490160

FINAL VALUES OF THETA (STANDARD ERROR) AFTER 626 ITERATIONS
B0 = -2.890781444440275 (.0857175370071179)
N = 7168.601091680866150 (420.9670461997142333)

CAPTURE PROB. (P) =
.052611155 .052611155 .052611155 .052611155 .052611155 .052611155 .052611155 .052611155 .052611155 .052611155
.052611155 .052611155 .052611155

THETA VARIANCE-COVARIANCE MATRIX

.007347496162	-35.360610133262
-35.360610133262	177213.253986112337

****
** LOG LIKELIHOOD FUNCTION = 16820.500066597454900 **
****

****
** LIKELIHOOD RATIO TEST **
** 3 PARAMETER MODEL VS. 2 PARAMETER MODEL **
** CHI-SQUARE VALUE = 449.2350 **
** PROB. = .0000 WITH 1 DEGREE OF FREEDOM **
****

GOODNESS OF FIT TEST FOR THE 2 PARAMETER MODEL

I	U(I)	E(I)	CHI-SQUARE
1	604.	377.15	136.44931
2	495.	357.31	53.06258
3	282.	338.51	9.43299
4	207.	320.70	40.31001
5	119.	303.83	112.43513
6	156.	287.84	60.38811
7	132.	272.70	72.59277
8	254.	258.35	.07328
9	299.	244.76	12.02042
10	325.	231.88	37.39404
11	247.	219.68	3.39700
12	276.	208.12	22.13619
13	222.	197.17	3.12560

TOTAL CHI SQUARE = 562.82
DEGREES OF FREEDOM= 11.
OVERALL PROB. = .0000

END OF PROGRAM