Part 7. Application of Theory

7.1. Introduction

Several examples of real data taken from various experiments in the literature are analyzed here to give insight and understanding into the statistical theory and the interpretation of experimental results. The examples here relate to studies where recaptures are made over time (days, weeks, or years), whereas Parts 1 and 2 dealt mostly with recaptures recorded spatially at a sequence of downstream dams. The duality between temporal and spatial sampling was noted in Section 1.1.2. Examination of this material should provide the reader experience with program RELEASE and a better perspective of the extent of this class of experiments beyond the fisheries emphasis provided in Parts 1 and 2. Lastly, several subtle points are made; and we mention some extended approaches.

The material presented here is not a critical reanalysis; our results are not in conflict with the original work of the investigators. We refer the reader to the original publications for biological conclusions. We can only briefly describe the more important features of these experiments. Again, we encourage the reader to consult the original sources for detailed information on methods and results.

We urge meticulous care in the processing of data before analysis by the statistical methods presented here. Errors in data entry are always a concern. Computer scatter plots of the m_{ti} versus m_{ci} often reveal mistakes that can be corrected or outliers that are cause for concern. Generally we recommend that the data be recorded and entered into program RELEASE in the form of a capture history matrix because hand summarization of the data into an m_{ij} matrix is error prone and does not allow full testing of model assumptions.

Many more-advanced models can be considered for well-designed studies supported by adequate sample sizes and replication. We urge the reader to become acquainted with the capabilities of program SURVIV (White 1983). Program RELEASE will prepare an input file for program SURVIV. Modification of this input file allows estimation with more complex models.

A series of contingency tables can be made to check for homogeneity and to suggest subsets to be pooled in the final analysis. We refer the reader to Olson and Kaczynski (unpublished report, 1980) for an example of the extensive testing that should be done before a final analysis is attempted. As the investigator explores the experimental data, thought should be given to what assumptions seem reasonable. Are chronic effects of the treatment likely? Has the recapture effort across sampling occasions been nearly constant? (If so, perhaps all the recapture rates can be modeled as a constant.) Is the treatment likely to affect the recapture rates? These considerations are important in modeling the data, selecting an appropriate

model, and interpreting the output of program RELEASE.

7.2. Lead-Dosing Experiments on Mallards

Bellrose (1959) conducted a series of experiments in the early 1950s to estimate the mortality caused by ingestion of lead shot by mallards *Anas platyrhynchos*. He used a sequence of treatments with one, two, and four lead pellets, each with a separate control (zero pellets). The experiments consisted of trapping and banding mallards during late fall and winter; every other duck was a treatment bird and the others were controls. Treatment birds were dosed with lead pellets and immediately released with control birds. The data came from recoveries of birds shot during annual sport-hunting seasons and, therefore, represent the first capture history protocol (see Table 7.1). The underlying parameterization is not exactly the same as that described in Chapter 2.2, as band recoveries during the year of banding were ignored in this example.

Estimates of treatment effect are made by using each treatment group versus its control group, assuming that $H_{1\phi}$ is the correct model. Here, it seems biologically reasonable to consider $H_{1\phi}$ as it corresponds to the direct effect

$$\phi_{t1} = S\phi_{c1}$$

and

Table 7.1. – Summary of Bellrose's (1959) data on lead-dosed mallards. The data represent recoveries shot by sport hunters and fall under the first capture history protocol.

Experiment number	Banding year	Age at banding	Number pellets	Number banded	Number recovered in year[a]		
					2	3	4
1	1949	Adult	1	559	52	36	22
			0	560	56	44	24
2	1950	Adult	2	277	12	9	4
			0	278	44	22	7
3	1951	Adult	4	284	13	7	2
			0	396	30	16	7
4	1950	Young	2	115	14	10	4
			0	111	12	12	2
5	1951	Young	4	220	14	8	2
			0	207	28	14	8

[a]Numbers in this table were provided by F. C. Bellrose and differ slightly from those in Bellrose (1959).

$$\lambda_{t1} = S\lambda_{c1}.$$

Alternatively, if the harvest rate of dosed mallards were to increase and the annual survival rate to decrease, one would have

$$p_{ti} > p_{ci}$$

and

$$\phi_{ti} < \phi_{ci},$$

and these conditions might result in $\lambda_{t1} \sim \lambda_{c1}$. If these conditions occur, the treatment effect, H_0 versus $H_{1\phi}$, would be difficult to detect.

If all the experiments for adults (or young) had been conducted the same year, all the data from the control groups could have been pooled for increased statistical efficiency. Estimates of treatment effect are given in Table 7.2. These estimates appear to show that lead pellets decreased survival. The tests of the null hypothesis of no treatment effect ($S = 1$) were rejected, except in experiment 4. A crude test of the null hypothesis can be computed by pooling the data for all the treatment groups versus all the control groups. Pooling all recoveries over the 3 years and five experiments results in the 2×2 table

t	209	1,246
c	326	1,226

which results in a chi-square value of 22.64, 1 df, $P = <0.001$, providing strong evidence of a treatment effect. This test is recommended to provide the investigator insight into the experimental results and to give background familiarity with the data.

Further insight is provided by examining the 2×2 contingency table for each treatment level (one, two, or four pellets) versus its control (zero pellets). Data are pooled over age for two- and four-pellet treatments as an example. In each treatment, the form of the table is from TEST 1.*R*1.

r_{t1}	$R_{t1} - r_{t1}$
r_{c1}	$R_{c1} - r_{c1}$

Table 7.2. – Summary of results of the survival experiments conducted by Bellrose (1959).

Experiment number	Dose	Survival $\hat{S}$	se($\hat{S}$)	H_0 versus $H'_{3\phi}$ χ^2 (3 df)	P	$H_{1\phi}$ versus $H'_{3\phi}$ χ^2 (2 df)	P
1	1	0.889	0.1036	1.23	0.747	0.20	0.906
2	2	0.344	0.0741	29.69	<0.001	1.35	0.509
3	4	0.579	0.1397	5.61	0.132	0.24	0.880
4	2	1.039	0.2470	0.95	0.812	0.93	0.629
5	4	0.452	0.1033	13.95	0.003	0.89	0.640

The contingency tables and test statistics are available through program RELEASE:

Dose	v	Contingency table	χ^2	df	P (one-tailed)
1	t	110 449			
	c	124 436	1.03	1	0.155
2	t	53 339			
	c	99 290	17.73	1	<0.001
4	t	46 458			
	c	103 500	14.91	1	<0.001

From this summary, it seems clear that a treatment effect is indicated; however, the treatment effect may be on survival or recapture rates, or both.

Estimates of S may be biased if $H_{1\phi}$ is not the correct model. This condition may result when the effect of the treatment is chronic rather than acute. The sum of TESTs 1.*Ti* represents a goodness of fit test to model $H_{1\phi}$, and these test results (Table 7.2) fail to provide evidence of chronic effects. These goodness of fit tests have low power in this example due to the small values of the m_{v3} and m_{v4} (see Table 7.1). If chronic effects of the treatment are reflected in ϕ_{t2}, ϕ_{t3}, p_{t2}, p_{t3}, or p_{t4}, then $\hat{S} = \hat{\phi}_{t1}/\hat{\phi}_{c1}$ under model $H_{1\phi}$ may be a poor measure of the magnitude of the treatment effect.

In experiments such as Bellrose's where a sequence of treatments was involved, one might consider estimating a function relating treatment survival to the number of pellets. We would then expect survival to decrease as the number of lead pellets increases. A generalized logistic model (Cox 1970) is often useful in modeling survival data (see discussion in Part 8).

The information above indicated a significant treatment effect, but not necessarily a direct, acute effect, $S = \phi_{t1}/\phi_{c1}$. Thus, it is reasonable to consider $S \doteq \lambda_{t1}/\lambda_{c1}$, where

$$\lambda_{v1} = E\left(\frac{r_{v1}}{R_{v1}}\right).$$

This formulation ignores the fact that a few birds were recovered during and shortly after initial banding and release. We modeled treatment effect as a function of dosage (D = 1, 2, 4) and age (A = 0, 1), ignoring year effects, as

$$E(\hat{S}) = \left[1 + e^{-[a_1 + a_2(D) + a_3(A)]}\right]^{-1}.$$

We estimated the parameters a_i, using the ML approach suggested by Jennrich and Moore (1975) as implemented in program BMDPLR (Dixon 1983). In this example $\hat{a}_1$, $\hat{a}_2$, and $\hat{a}_3$ are 2.23, -0.57, and -0.28, respectively. The respectively estimated standard errors are 0.15, 0.05, and 0.07. For a preliminary analysis, the variables can be transformed and run with a linear regression program such as

$$\ln\left[\frac{\hat{S}_i}{1 - \hat{S}_i}\right] = a_1 + a_2(D_i) + a_3(A_i).$$

We do not recommend this analysis as a final procedure, but merely as one that enables quick insight into the uses of this general modeling approach. Although both dose and age are significant in this logistic model, the model is crude because the data are too sparse to support much modeling (temporal effects are ignored as there are only five treatments involving two age-classes). We use this logistic analysis only as an example of what can be done when sequences of treatments are used.

Finally, we point out that the five treatment survival estimates in Bellrose's studies were independent (i.e., each treatment had a paired control). In other studies, several treatments may have used a common control. The $\hat{S}_i$ will then have sampling correlations that lead to a weighted analysis where the weighting matrix is not diagonal.

7.3. Lead-Dosing Study of Northern Pintails

Deuel (1985) reported on a study that measured indirect mortality from ingested lead shot in northern pintails *Anas acuta*. The ducks were caught in baited traps at seven areas in California in the winter months in early 1979. Birds were banded with aluminum leg bands,

and every other bird received a treatment consisting of two #5 lead shot pellets put directly into the crops through a plastic tube inserted into the esophagus. Otherwise, all birds were handled alike and released. In subsequent hunting seasons, some of the banded birds taken by hunters were reported to the Bird Banding Office of the U.S. Fish and Wildlife Service.

We use the data from five of Deuel's (1985) areas for male pintails to illustrate the methodology presented in this monograph (relatively fewer females were banded and released, and three of the areas were poorly represented by banding, even for males). The data we use (Table 7.3) fall under the first capture history protocol. The numbers and the capture probability are both fairly small (e.g., $\hat{p}_2 \sim 0.04$).

Data collected under the first capture history protocol do not permit intensive tests of model assumptions, nor are the capture and survival rates separately estimable. The only estimable parameter of concern is S, the treatment effect. Estimates of this parameter under model $H_{1\phi}$ and results of TEST 1, the test for a treatment effect, are provided Table 7.4 for each of the five areas and the pooled data. There is no evidence of a significant treatment effect for any of the five areas (Table 7.4). A pooled test statistic, computed by summing the five chi-squared values and their degrees of freedom, yields $\chi^2 = 17.9$, 20 df, and $P = 0.60$. Thus, these data fit model H_0 quite well.

Table 7.3. – Summary of the experimental lead dosing study of male northern pintails banded in California in 1979 (from Deuel 1985). These data fall under the first capture history protocol.

Banding area	Group	Number released	Number recovered in year j					
			$j = 2$	3	4	5	6	Total
Mendota	t	930	27	29	12	9	8	85
	c	932	41	16	12	9	8	86
S. Grasslands	t	759	31	16	6	12	7	72
	c	759	28	16	8	9	8	69
Yolo Bypass	t	558	17	15	8	9	1	50
	c	562	20	8	4	7	4	43
Gray Lodge	t	1,354	45	26	13	20	10	114
	c	1,334	43	30	19	24	11	127
Delevan/	t	712	23	17	5	7	7	59
Colusa	c	783	19	14	6	11	7	57
Pooled	t	4,313	143	103	44	57	33	380
	c	4,370	151	84	49	60	38	382

Table 7.4. – Estimates of treatment survival under model $H_{1\phi}$ and the results of TEST 1 for treatment effect for the lead dosing study of male northern pintails banded in California in 1979 (from Deuel 1985).

Banding area	Estimated survival $\hat{S}$	Standard error $\hat{se}(\hat{S})$	Test of H_0 versus $H'_{5\phi}$		
			χ^2	df	P
Mendota	0.990	0.1444	6.67	5	0.246
S. Grasslands	1.043	0.1674	0.94	5	0.967
Yolo Bypass	1.171	0.2333	5.90	5	0.316
Gray Lodge	0.884	0.1089	2.17	5	0.825
Delevan/ Colusa	1.138	0.2030	2.15	5	0.828
Pooled data	1.008	0.0697	2.85	5	0.722

Further extending this example, we note that the five estimates of treatment survival computed under model $H_{1\phi}$ are all near 1.0, considering the size of their standard errors. The estimate of 1.008 for treatment survival computed by pooling the data over the five areas indicates a lack of treatment effect due to dosing with two lead shot pellets. Alternatively, TEST 1.R1 of the null hypothesis that $S = 1$ (H_0 versus $H_{1\phi}$) can be made for each area, and the results of the individual TEST 1.R1 can then be pooled for an overall test:

Area	χ^2	df	P
1	0.00	1	0.948
2	0.07	1	0.791
3	0.63	1	0.427
4	1.00	1	0.318
5	0.53	1	0.467
Total	2.23	5	0.816

Again, any indication of mortality due to the lead treatment appears to be lacking. Additional details on this study were given by Deuel (1985). Although his approach differed from that presented here, his conclusions were similar. Readers are encouraged to analyze these data by using program RELEASE to gain additional insights and familiarity with the various models and tests.

Deuel's (1985) data can be used to illustrate several other technical issues because the experiment was replicated over five areas. First we discuss estimates of treatment survival over the entire experiment. In view of the replicated nature of the study, the individual estimates of S_i (by area i) and their standard errors, and the results of testing the null hypothesis that $S = 1$, it is logical to pool the raw data and proceed to estimate S. This method yields an estimate of 1.008, with a theoretical $\hat{se}(\hat{S}) = 0.0697$. Under the null hypothesis, this outcome is satisfactory. However, the estimated standard error may be somewhat poor unless the

$\text{se}(\hat{S}_i)$ is equal for all i.

Alternatively, a simple average of the five estimates gives 1.046, with an empirical $\hat{\text{se}}(\hat{S})$ = 0.0517. It can be argued that this procedure is poor because some of the estimates have larger standard errors than others (observed range = 0.109 to 0.233), suggesting that some estimates should be given more weight (and therefore requiring a weighted average). If we assume all five estimates are of the same parameter, statistical theory states that the proper, optimal weight (w) is $\text{var}(\hat{S})^{-1}$, i.e., the inverse of the true sampling variance. We can estimate these weights as

$$w_i = \left\{(\hat{S}_i^2)\left[\frac{1}{r_{t1}} - \frac{1}{R_{t1}} + \frac{1}{r_{c1}} - \frac{1}{R_{c1}}\right]\right\}^{-1}$$

and

$$\overline{S}_{wt} = \frac{\sum w_i \hat{S}_i}{\sum w_i}$$

$$= 0.989 \text{ or } 99\%.$$

This procedure is flawed because the estimate of S_i and its estimated sampling variance are positively correlated. The result causes estimates that are too low to have an estimated sampling variance that is too low and vice versa. Thus, low estimates receive a weight that is too large, which causes estimates of the weighted average to be biased low. The reason for the estimate being related to its own variance estimate can be seen by noting that the variance estimator contains the term $(\hat{S}_i)^2$ (see equation above).

A reasonable alternative is to weight the individual estimates by the final term of $\hat{\text{var}}(\hat{S}_i)$, which is a measure of "sample size." Actually, this final term is the (large-sample) variance of $\ln(\hat{S})$ (see Part 3). Here the weights are defined as

$$w_i = \left[\frac{1}{r_{t1}} - \frac{1}{R_{t1}} + \frac{1}{r_{c1}} - \frac{1}{R_{c1}}\right]^{-1},$$

and the weighted mean is computed as

$$\overline{S}_{wt} = \frac{\sum w_i \hat{S}_i}{\sum w_i}.$$

When the pintail data are used, the weights are 47.1, 38.3, 25.2, 66.0, and 31.4. This procedure yields a weighted mean of 1.015, close to the value obtained by pooling the data. Corresponding to the estimator of the weighted mean is an empirical estimator of its sampling variance,

$$\text{var}(\overline{S}_{wt}) = \frac{\sum_{i=1}^{5} w_i(\hat{S}_i - \overline{S}_{wt})^2}{\left(\sum_{i=1}^{5} w_i\right)(5-1)}.$$

For the pintail data, $\hat{\text{var}}(\overline{S}_{wt}) = 0.00272$, or $\hat{\text{se}}(\overline{S}_{wt}) = 0.0522$.

Another subject of concern is the degree to which theoretical sampling variances from the model reflect the amount of variation in the experiment. Using the pintail data, we can compare the sampling variance based on the model with the variance computed empirically. The comparison below is based on the unweighted mean but the extension to weighted means is straightforward.

Maximum likelihood theory provides the five estimates, $\hat{S}_i$, and their associated sampling variances, $\hat{\text{var}}(\hat{S}_i)$. If we take

$$\overline{S} = \frac{1}{5} \sum_{i=1}^{5} (\hat{S}_i),$$

then the ML estimator of the sampling variance of this average is estimated by

$$\hat{\text{var}}(\overline{S}) = \left(\frac{1}{5}\right)^2 \sum_{i=1}^{5} \hat{\text{var}}(\hat{S}_i),$$

because the five data sets are independent. The $\hat{\text{se}}(\overline{S}) = [\hat{\text{var}}(\overline{S})]^{1/2} = 0.079$.

Alternatively, an empirical variance of $\overline{S}$ could be computed as

$$\hat{\text{var}}(\overline{S}) = \frac{1}{5(4)} \sum_{i=1}^{5} (\hat{S}_i - \overline{S})^2 .$$

Then, $\hat{\text{se}}(\overline{S}) = [\hat{\text{var}}(\overline{S})]^{1/2} = 0.052$, which is in reasonable agreement with the theoretical (model-based) estimate of 0.079. However, one could question whether the theoretical variance (or standard error) is too large. A useful guideline here is to compute

$$\frac{(n - 1)\ (\text{empirical variance } of\ \overline{S})}{(\text{theoretical variance } of\ \overline{S})} ,$$

which is asymptotically distributed as χ^2 with n - 1 df, where n is the number of survival rates being averaged. This procedure tests the null hypothesis that the theoretical variance and the empirical variance are equal. Thus,

$$\frac{(5 - 1)\ (0.052)^2}{(0.079)^2} = 1.73 .$$

However, a value of 1.73 for a chi-square variable with 4 df is not unusual (P = 0.785) and provides no evidence for rejecting the null hypothesis. Usually, one is more concerned over the possibility that the empirical variance is larger than the theoretical (i.e., model-based) variance.

One possibility is that the true treatment survival varies among the five areas. This variation leads to a variance component we call σ_a^2, in contrast to the sampling variance associated with each estimate. (Statisticians would write this latter term as var[$S_i \mid S_i$] to indicate that it is only sampling variation.) In this example, three estimates were >1.0; but the true rates cannot, of course, exceed 1.0, unless one naively believes that lead enhances survival in biological organisms. Nevertheless, a method of separating the variance components is illustrated. Variance component estimation can be difficult; here we use a procedure given by Anderson and Burnham (1976):

$$\hat{\sigma}_a^2 = \frac{1}{n-1}\sum_{i=1}^{n}(\hat{S}_i - \overline{S})^2 - \frac{1}{n}\left[\sum_{i=1}^{n}\hat{\text{var}}(\hat{S}_i \mid S_i)\right],$$

where $\hat{\sigma}_a^2$ is the estimator of the area-to-area population variance in S_i. The estimator of $\sigma_a{}^2$ is similar for the weighted approach

$$\hat{\sigma}_a^2 = \frac{1}{n-1}\frac{\sum_{i=1}^{n} w_i(\hat{S}_i - \overline{S}_{wt})^2}{\left(\sum_{i=1}^{n} w_i\right)} - \frac{1}{n}\left[\sum_{i=1}^{n}\hat{\text{var}}(\hat{S}_i \mid S_i)\right].$$

In this example, $\hat{\sigma}_a^2$ = -0.01777 (if weighting is used, this estimate is similar at -0.0176). Variance is a positive quantity; however, an estimate can be negative. Here the estimate is close to zero and we could conclude that we were unable to attribute any significant variation to an area-to-area component. This conclusion is logical given the proximity of the areas and the nature of the treatment and sampling program.

The pintail data can be separated into 10 replicates based on the last digit of the band number. Separation of data allows another assessment of the precision of the estimate of survival pooled over the five areas. These replicates can be considered as true replicates and are summarized in Table 7.5. The individual estimates of treatment survival (Table 7.6) yielded an unweighted average of 1.024. The standard error of this simple average can be computed in two ways. First the model-based estimates of the sampling variance of each of the 10 estimates can be used as

$$\hat{\text{var}}(\overline{S}) = \left(\frac{1}{10}\right)^2\left[\sum_{i=0}^{9}\text{var}(\hat{S}_i)\right] = 0.0053$$

and

$$\hat{\text{se}}(\overline{S}) = \sqrt{\hat{\text{var}}(\overline{S})} = 0.0729.$$

Second, the sampling variance of the mean can be computed empirically from the 10 independent estimates as

$$\hat{\text{var}}(\overline{S}) = \frac{1}{10(9)} \sum_{i=0}^{9} (S_i - \overline{S})^2 = 0.0063$$

and

$$\hat{\text{se}}(\overline{S}) = \sqrt{\hat{\text{var}}(\overline{S})} = 0.0794 .$$

These estimates of precision are in close agreement with each other (0.073 versus 0.079) and to the comparable estimate of 0.079 based on the theoretical variance and the empirical estimates computed for the five areas (0.052).

Table 7.5. – Summary of the experimental lead dosing study of male northern pintails banded in California in 1979 (Deuel 1985). The data are pooled over the five areas and segregated into 10 replicates, based on the last digit of their band number; $R_{t1} = 431$ and $R_{c1} = 437$ in each replicate.

Replicate number	Group	Number recovered in year j					Total
		$j = 2$	3	4	5	6	
0	t	18	14	2	10	4	48
	c	13	9	1	6	3	32
1	t	13	5	4	7	1	30
	c	19	5	5	8	4	41
2	t	11	11	6	3	0	31
	c	15	7	9	7	4	42
3	t	12	12	9	6	6	45
	c	14	10	6	7	2	39
4	t	17	12	6	7	4	46
	c	18	9	4	5	6	42
5	t	9	9	2	5	3	28
	c	15	7	6	7	1	36
6	t	17	11	4	7	1	40
	c	18	7	8	5	5	43
7	t	16	10	3	3	5	37
	c	9	9	5	4	5	32
8	t	11	10	4	4	5	34
	c	12	11	4	9	4	40
9	t	19	9	4	5	4	41
	c	18	10	1	2	4	35

Table 7.6. – Estimates of treatment survival and associated statistics under $H_{1\phi}$ for the pintail dosing experiment (Table 7.5). Replicates are derived by partitioning the data on the basis of the last digit in the band number.

Replicate	Estimated survival $\hat{S}$	Standard error $\hat{se}(\hat{S})$	Null hypothesis $S = 1$[a]		
			χ^2	df	P
0	1.521	0.3314	3.95	5	0.557
1	0.742	0.1710	3.14	5	0.678
2	0.748	0.1698	8.45	5	0.133
3	1.170	0.2433	3.20	5	0.670
4	1.110	0.2247	1.68	5	0.891
5	0.789	0.1914	5.10	5	0.404
6	0.943	0.1970	5.27	5	0.383
7	1.172	0.2716	2.78	5	0.734
8	0.862	0.1923	2.13	5	0.830
9	1.188	0.2612	3.42	5	0.636
Mean or total	1.025		39.12	50	0.867

[a]TEST 1.

7.4. Pesticide Dosing of Starlings

Stromborg et al. (in press) studied postfledgling survival of European starlings *Sturnus vulgaris* deliberately exposed to an organophosphate pesticide. This novel study of the effect of a pesticide under field conditions represents another example of other types of experiments that fall under the general methodology developed here. In the starling experiment, relatively small numbers of birds were released ($R_{t1} = 60$; $R_{c1} = 61$), but capture probabilities were high ($\overline{p} = 0.78$). Inasmuch as the recaptures were resightings of uniquely marked birds, the experiment falls under what we have termed the complete capture history protocol (see Nichols et al. 1981; Sandland and Kirkwood 1981; Buckland et al. 1983; and Clobert et al. 1985 for similar studies, but without treatments). The study was conducted on about 2,000 hectares of the Patuxent Wildlife Research Center near Laurel, Maryland.

During summer 1984, investigators set out nest boxes to attract starlings. Boxes were checked frequently during the nesting period to determine the date of hatching. All nestlings were banded 16 days after hatching, and half the birds, chosen at random, were given an oral dosage of organophosphate pesticide mixed in corn oil. Control birds were given pure corn oil under similar conditions. Two days later, the surviving birds were tagged with individually numbered wing tags (dead birds were tallied as "direct" mortality). The tags were made of orange or red vinyl-like material, cut in a pear shape and measuring approximately 35 × 40 mm. Letters or numerals about 20 mm high were painted on each tag with flat black paint. Tagged birds were sighted and their tag numbers were recorded over six sampling periods. (Further details are given by Stromborg et al., in press.)

Data from the study are presented in Table 7.7, taken directly from the output of program RELEASE in the form of a reduced m-array and its associated summary statistics. Stromborg et al. (in press) failed to find any significant treatment effect, and our analysis is in agreement with their results. Results for TEST 1 are summarized in the series of 2×2 contingency tables given in Table 7.8. Pooling of the chi-square values and their degrees of freedom results in an overall test of the null hypothesis of equality for all survival and capture probabilities between the treatment and control groups. This procedure yields $\chi^2 = 8.26$, 9 df, and $P = 0.508$; thus, there is no reason to suspect the validity of the null hypothesis. Either the pesticide had no effect on the survival rate of fledged starlings or the effects were too small to be detected in this limited experiment.

The estimated treatment survival under model $H_{1\phi}$ is 0.906 (standard error, 0.1036), which is not significantly different from 1.0 (in agreement with results of TEST 1.R1). If $S <$ 1, a larger experiment is needed to provide a suitable measure of the treatment effect. Alternatively, a higher dose level or sequence of doses might be effective with the same sample sizes.

If the treatment affected survival throughout the experiment, a comparison of the average survival rates of the two groups would provide an indication of the extent of that effect. However, averaging the $\hat{\phi}_i$ over the four estimates available yielded 0.862 and 0.880 for treatment and control birds, respectively. Again, no indication of treatment-related mortality is shown. Readers are encouraged to run the data in Table 7.7 through program RELEASE and interpret the full output.

Table 7.7. – Data under the complete capture history protocol for survival studies of marked starlings dosed with pesticide by Stromborg et al. (in press).

	R_{ti} or	Recaptures, m_{tij}					r_{ti} or
i	R_{ci}	$j = 2$	3	4	5	6	r_{ci}
				Treatment group			
1	60	24	6	9	2	0	41
2	24		22	1	0	0	23
3	28			24	0	0	24
4	34				30	0	30
5	32					21	21
m_{tj}		24	28	34	32	21	
z_{tj}		17	12	2	0	0	
				Control group			
1	61	22	13	9	2	0	46
2	22		18	1	0	0	19
3	31			30	0	0	30
4	40				33	1	34
5	35					28	28
m_{cj}		22	31	40	35	29	
z_{cj}		24	12	2	1	0	

Table 7.8. – A series of 2 × 2 contingency tables and associated information related to the null hypothesis that survival and capture (sighting) probabilities are equal for treatment and control groups of starlings.

Test 1 component	Contingency table		χ^2	df	P
1.*R*5	21	11			
	28	7	1.76	1	0.185
1.*T*5	32	0			
	35	1	0.90	1	0.342
1.*R*4	30	4			
	34	6	0.16	1	0.685
1.*T*4	34	2			
	40	2	0.03	1	0.874
1.*R*3	24	4			
	30	1	2.32	1	0.128
1.*T*3	28	12			
	31	12	0.04	1	0.834
1.*R*2	23	1			
	19	3	1.30	1	0.255
1.*T*2	24	17			
	22	24	1.00	1	0.318
1.*R*1	41	19			
	46	15	0.75	1	0.387
TEST 1			8.26	9	0.508

We used program RELEASE to simulate 1,000 replications of this study to investigate the performance of the statistical theory for several experiments similar to the starling study. The results of these studies are used to illustrate a number of points and to allow additional insight into the effect of a pesticide on young starlings. In general, we took the estimates of parameters computed from the starling study as parameters for the simulations. Thus, $\phi_2 = 0.94$, $\phi_3 = 0.92$, $\phi_4 = 0.87$, $\phi_5 = 0.87$ and $p_2 = 0.5$, $p_3 = 0.7$, $p_4 = 0.95$, $p_5 = p_6 = 0.98$ for both groups. We chose $\phi_{t1} = 0.71$ and $\phi_{c1} = 0.79$, giving $S = 0.899$. The true model was then $H_{1\phi}$ and data sets were generated under this model. The total number released in each group was studied at 60, 100, 200, and 400 individuals.

The statistical procedures performed well on the average, for all sample sizes studied (Table 7.9). In addition, the empirical variance of the 1,000 estimates is in close agreement with the average variance derived from the model and the ML method. The above results indicate that the asymptotic theory does well when small numbers of animals are released if the recapture rates are high. We recommend that simulations such as this one be performed before a study is conducted and during the analysis of experimental data. Simulations conducted during the design phase of an experiment allow appraisal of precision and sample size

(e.g., note how precision increases in the final column of Table 7.9 as sample size increases). Also, the power of various tests can be evaluated. The standard error of $\hat{S}$ is relatively large when sample sizes are small.

Alternatively, the use of the option EXPECT in program RELEASE will provide very reliable insight into the properties of estimators and tests. The theory for this is given in Chapter 3.6. The use of the EXPECT option produces results in a few seconds of computer time whereas simulation times may often exceed 3 hours.

Table 7.9. – Monte Carlo results for four parameters and four sample sizes for experiments similar to the study of the effect of a pesticide on starlings by Stromborg et al. (in press). One thousand replications were generated with $S = 0.899$ and estimates were made under model $H_{1\phi}$ (the true model).

Number released $R_{t1} = R_{c1}$	Parameter	Average of estimates[a]	Standard error
60	$p_2 = 0.5$	0.501	0.055
	$\phi_{t1} = 0.71$	0.708	0.063
	$\phi_{c1} = 0.79$	0.793	0.060
	$S = 0.899$	0.898	0.102
100	$p_2 = 0.5$	0.500	0.043
	$\phi_{t1} = 0.71$	0.712	0.049
	$\phi_{c1} = 0.79$	0.790	0.048
	$S = 0.899$	0.903	0.080
200	$p_2 = 0.5$	0.501	0.031
	$\phi_{t1} = 0.71$	0.710	0.035
	$\phi_{c1} = 0.79$	0.792	0.034
	$S = 0.899$	0.899	0.055
400	$p_2 = 0.5$	0.501	0.022
	$\phi_{t1} = 0.71$	0.709	0.025
	$\phi_{c1} = 0.79$	0.791	0.024
	$S = 0.899$	0.896	0.040

[a]An estimate of the expected value of the estimator.

Table 7.10. – Power of the test (TEST 1.R1) of model $H_{1\phi}$ versus H_0 for five significance levels and four sample sizes in Monte Carlo studies with parameters similar to those estimated from the starling data of Stromborg et al. (in press). Data were simulated under model $H_{1\phi}$ where the survival rates averaged 0.9 and the recapture rates averaged 0.8. The treatment survival $S = \phi_{t1}/\phi_{c1}$.

Treatment survival	Number released $R_{t1} = R_{c1}$	Significance level (α)				
		0.01	0.05	0.10	0.20	0.50
$S = 0.9$ (small effect)	60	0.05	0.17[a]	0.24	0.38	0.65
	100	0.09	0.22	0.33	0.45	0.72
	200	0.20	0.40	0.53	0.68	0.87
	400	0.47	0.70	0.80	0.89	0.97
$S = 0.8$ (larger effect)	60	0.22	0.44	0.58	0.72	0.88
	100	0.42	0.68	0.79	0.87	0.96
	200	0.80	0.93	0.96	0.99	1.0
	400	0.98	1.0	1.0	1.0	1.0

[a]Comparable estimates of test power using the theory given in Chapter 3.6 are 0.16, 0.24, 0.42, 0.70, 0.47, 0.68, 0.93, and 1.00, respectively, for this column.

Simulated results on the power of the test of $H_{1\phi}$ versus H_0 are shown in Table 7.10. The top half of the table relates to the simulations in Table 7.9. The rest of the table relates to a larger treatment effect, where $S = \phi_{t1}/\phi_{c1} = 0.63/0.79 = 0.797$. The power of finding a significant treatment effect, at the $\alpha = 0.05$ level of significance where $S = 0.9$, is only about 0.17 for the starling experiment. Power increases with sample size and the magnitude of the treatment effect. The ideal experiment would consist of a larger treatment effect (i.e., low S), large samples released, high capture probabilities, many sampling sites or occasions, adequate replication, and the use of unique marks. The use of ordered treatments also has advantages (see Section 8.2.1).

Data collected under the complete capture history protocol have the advantage that one can make many tests of model assumptions, select an appropriate estimator, and make proper estimates of parameters. As noted by Stromborg et al. (in press), the goodness of fit tests indicate a poor fit of the data to the model. This matter is potentially serious. Stromberg et al. (in press) simulated the lack of fit and concluded that, in their case, the results were little affected. In this example, no treatment effect could be shown; thus, model H_0 was selected. If a significant treatment effect were to exist, the investigator would have a choice of models: $H_{1\phi}$, H_{2p}, $H_{2\phi}$, H_{3p}, ..., $H_{k\text{-}1,\phi}$. In general, however, only models $H_{1\phi}$ and H_{2p} have relatively good precision unless the capture probabilities are high. For the starling data, the full output from program RELEASE shows good precision for all estimators under all models because the capture (sighting) probabilities are high. If capture probabilities are low (say < 0.1), then good precision for S can be expected only when at least $\phi_{tj} = \phi_{cj}$ for $j = 2, 3, ..., k$ and $p_{tj} = p_{cj}$ for $j = 3, 4, ..., k$. This same situation is also true for scheme A in the partial capture history protocol. Under scheme B one can test $\phi_{t1} = \phi_{c1}$ and $p_{t2} = p_{c2}$ and also test that the other parameters are the same by treatment and control group. With scheme B data, one can estimate the relevant ϕ_{t1} and ϕ_{c1} and the treatment survival rate S can be estimated under models $H_{1\phi}$ and H_{2p}; however, ϕ_{ti} for $i > 1$ cannot be estimated under any model.

7.5. Partitioning Lazuli Bunting Data

Allen W. Stokes banded Lazuli buntings *Passerina amoena* in his yard in Logan, Utah, during winters from 1974 to 1980. Recaptures were recorded each year for 7 consecutive years, and his data (personal communication), summarized as an m-array, are shown in Table 7.11. The result of the TEST 2 goodness of fit tests is summarized below:

TEST	χ^2	df	P
2.*C*2	1.74	1	0.187
2.*C*3	1.68	1	0.194
2.*C*4	0.32	1	0.571
2.*C*5	2.29	2	0.130
2.*C*6	0.21	1	0.643
Total	6.25	5	0.282

(with only an m-array such as in Table 7.11, only TEST 2 goodness of fit can be computed).

Table 7.11. – The m-array for Lazuli buntings banded during winters of 1974-1980 by Allen W. Stokes in Logan, Utah (personal communication).

		Recaptures, $m_{i,j}$							
i	R_i	1974	1975	1976	1977	1978	1979	1980	r_i
1	168	31	4	0	1	0	0	0	36
2	398		19	14	4	1	0	1	39
3	88			20	3	1	0	0	24
4	264				41	6	3	1	51
5	304					58	13	1	72
6	322						67	9	76
7	323							76	76
m_j		31	23	34	49	66	83	88	
z_j		5	21	11	13	19	12	0	

The above test results, taken from the output of program RELEASE, indicate a satisfactory fit of the Jolly-Seber model (model $H_{7\phi}$) to the data. The estimates of the recapture and survival rate parameters are given in Table 7.12. The average annual survival is 0.399. At this point, one might be satisfied with the model and proceed to make inferences about this banded population. A further analysis, however, raises many questions.

We will use these Lazuli bunting data to illustrate several testing, model-building, and model selection concepts. In particular, we use PROC SURVIV, an option in program RELEASE, to generate input code to program SURVIV for the bunting data. The input file is then analyzed with program SURVIV, which allows extended model building, testing, and estimation. In this example, we do not claim to reach a completely satisfactory endpoint; rather, these data are used to show a path that might be followed in the analysis of a set of real data where complications arise.

Table 7.12. – Estimates of annual survival (ϕ_i) and recapture (p_i) rates for the Lazuli bunting data under the Jolly-Seber model (model $H_{7\phi}$), output from program RELEASE.

Maximum Likelihood Estimates under Model H7Phi

Parameter	Estimate	Standard Error	95% Confidence Intervals Lower	Upper
		Estimates for Group 1		
Phi(1)	.488248	.142347	.209248	.767248
Phi(2)	.222704	.049248	.126178	.319231
Phi(3)	.551159	.128118	.300047	.802270
Phi(4)	.323701	.054803	.216288	.431114
Phi(5)	.408204	.058924	.292714	.523695
Phi(6)	.332919	.044263	.246163	.419675
p(2)	.377931	.118809	.145065	.610797
p(3)	.230000	.061706	.109056	.350944
p(4)	.373868	.086373	.204577	.543159
p(5)	.471658	.081868	.311196	.632119
p(6)	.450512	.069064	.315146	.585878
p(7)	.619403	.076551	.469364	.769442
Phi(7)p(8)	.235294	.023602	.189034	.281554

The full m-array is shown in Table 7.13 along with some associated statistics. TEST 3 can be made from these further partitions of the data. The results are summarized below in two components; first for TEST 3.*Sm*:

TEST	χ^2	df	P
3.*Sm*2	1.36	1	0.243
3.*Sm*3	1.64	1	0.200
3.*Sm*4	0.28	1	0.597
3.*Sm*5	6.68	1	0.010
3.*Sm*6	1.27	1	0.260
Subtotal	11.25	5	0.047

This component of TEST 3 shows some lack of fit but this is nearly all from release 5 (TEST 3.*Sm*5). The second major component of TEST 3 is a contingency table where columns are r_i and $R_i - r_i$ while rows are newly caught and released birds, and previously captured and released birds. The results of this test component (TEST 3.*SRi*) are summarized below:

TEST	χ^2	df	P
3.*SR*3	25.09	1	<0.001
3.*SR*4	13.43	1	<0.001
3.*SR*5	33.48	1	<0.001
3.*SR*6	10.14	1	0.006
3.*SR*7	24.58	1	<0.001
3.*SR*8	21.89	1	<0.001
Subtotal	128.62	6	<0.001

The sum of these two test components (TEST 3) is $\chi^2 = 139.87$, 11 df, $P < 0.001$ and indicates a serious lack of fit to the Jolly-Seber model. It is clear that the banded population consists of a mixture of birds that are probably never seen again after banding and birds that are commonly recaptured. The new, unbanded birds that enter the study population are different from those banded birds already in the marked study population. The birds that return and are recaptured fit the Jolly-Seber model fairly well. The data in Table 7.11 can be partitioned into two m-arrays: new captures and previous captures (Table 7.14). If we use the convention $v = 1$ for new birds to be released and $v = 2$ for previously banded birds already in the population at year j, we can compute the test H_0 versus $H_{7\phi}$ to examine the homogeneity of the two data sets. Note that $R_{21} = 0$: no releases on occasion 1 for group $v = 2$. This test can also be made using program RELEASE and results in $\chi^2 = 130.26$, 11 df, $P < 0.001$. This partitioning of the data can be done in other applications to test assumptions about males versus females, young versus adult, birds with versus without neck collars, etc. In each case, the relevant test is H_0 versus $H_{k\text{-}1,\phi}$.

Table 7.13. – The full m-array for the Lazuli bunting data. The total number of initial captures (r_i) and the number of birds never recaptured ($R_i - r_i$) are also shown.

	Occasion								
1	2	3	4	5	6	7	8	r_i	$R_i - r_i$
168	31	4	0	1	0	0	0	36	132
{11}	31	7	4	0	0	0	0	11	20
{01}	367	12	10	4	1	0	1	28	339
	{101}	4	1	0	1	0	0	2	2
(release 3)	{111}	7	5	0	0	0	0	5	2
	{011}	12	6	0	0	0	0	6	6
	{001}	65	8	3	0	0	0	11	54
		{1101}	4	4	0	0	0	4	0
		{0101}	10	3	1	0	0	4	6
(release 4)		{1011}	1	1	0	0	0	1	0
		{1111}	5	3	0	0	0	3	2
		{0111}	6	1	1	0	0	2	4
		{0011}	8	4	1	0	0	5	3
		{0001}	230	25	3	3	1	32	198
			{10001}	1	1	0	0	1	0
			{01001}	4	2	0	0	2	2
			{00101}	3	1	0	0	1	2
			{11011}	4	1	0	0	1	3
			{01011}	3	2	0	0	2	1
(release 5)			{10111}	1	1	0	0	1	0
			{11111}	3	1	0	0	1	2
			{01111}	1	0	0	0	0	1
			{00111}	4	2	0	0	2	2
			{00011}	25	9	0	0	9	16
			{00001}	255	38	13	1	52	203
				{010001}	1	0	0	0	1
				{101001}	1	0	0	0	1
				{010101}	1	0	0	0	1
				{011101}	1	1	0	1	0
				{001101}	1	1	0	1	0
				{000101}	3	0	1	1	2
				{100011}	1	0	0	0	1
				{010011}	2	0	0	0	2
(release 6)				{001011}	1	0	0	0	1
				{110111}	1	1	0	1	0
				{010111}	2	1	0	1	1
				{101111}	1	0	0	0	1
				{111111}	1	1	0	1	0
				{001111}	2	1	1	2	0

Table 7.13 – Continued.

	{000111}	9	6	0	6	3
	{000011}	38	16	0	16	22
	{000001}	256	39	7	46	210
		{0001001}	3	0	0	3
		{0000101}	13	1	1	12
		{0111011}	1	0	0	1
		{0011011}	1	0	0	1
		{1101111}	1	1	1	0
		{0101111}	1	1	1	0
(release 7)		{1111111}	1	1	1	0
		{0011111}	1	1	1	0
		{0001111}	6	2	2	4
		{0000111}	16	11	11	5
		{0000011}	39	12	12	27
		{0000001}	240	46	46	194

Table 7.14. – Partitioned m-arrays for the Lazuli bunting data. The data for birds recaptured only once ($v = 1$) appear at the top, followed by birds recaptured after being rereleased ($V = 2$). The statistics m_j and z_j are also shown.

		Recaptures, $m_{i,j}$							
i	R_i	$j = 2$	3	4	5	6	7	8	r_i
1	168	31	4	0	1	0	0	0	36
2	367		12	10	4	1	0	1	28
3	65			8	3	0	0	0	11
4	230				25	3	3	1	32
5	255					38	13	1	52
6	256						39	7	46
7	240							46	46
m_j		31	16	18	33	42	55	56	
z_j		5	17	10	9	19	10	0	
2	31		7	4	0	0	0	0	11
3	23			12	0	1	0	0	13
4	34				16	3	0	0	19
5	49					20	0	0	20
6	66						28	2	30
7	83							30	30
m_j		0	7	16	16	24	28	32	
z_j		0	4	1	4	0	2	0	

With the two extracted data sets in Table 7.14, the following parameter estimates were computed using program RELEASE:

	$v = 1$		$v = 2$	
i	ϕ_{1i}	$\hat{p}_{1i}$	$\hat{\phi}_{2i}$	$\hat{p}_{2i}$
1	0.575			
2	0.270	0.321	0.454	
3	0.544	0.137	0.591	0.497
4	0.255	0.200	0.720	0.900
5	0.494	0.428	0.408	0.621
6	0.296	0.284	0.508	1.000
7		0.513		0.835
Average	0.406	0.314	0.536	0.771

Here, it appears that the average annual survival rates are somewhat similar ($\overline{\phi}_1 = 0.406$ versus $\overline{\phi}_2 = 0.536$) but the average capture rates are quite different ($\overline{p}_1 = 0.314$ versus $\overline{p}_2 = 0.771$). Although one might expect the estimated average survival for the pooled data (Table 7.12) to lie between 0.406 and 0.536, this is not the case (0.399). These results show the danger of using the Jolly-Seber model without careful review of the results from TEST 3, which can only be computed from the CH matrix or the full m-array.

The analysis of these data was extended using PROC SURVIV, an option in program RELEASE (see Part 9). The following summarizes the parameters and indexing for the bunting data.

Releases	$j =$	3	4	5	6	7	8
Group $V = 1$							
$R_{12}, ..., R_{17}$		ϕ_{12}	ϕ_{13}	ϕ_{14}	ϕ_{15}	ϕ_{16}	ϕ_{17}
		p_{13}	p_{14}	p_{15}	p_{16}	p_{17}	p_{18}
Group $V = 2$							
$R_{22}, ..., R_{27}$		ϕ_{22}	ϕ_{23}	ϕ_{24}	ϕ_{25}	ϕ_{26}	ϕ_{27}
		p_{23}	p_{24}	p_{25}	p_{26}	p_{27}	p_{28}

PROC SURVIV was used to generate a computer file for analysis by program SURVIV (White 1983). A listing of this file is shown in Table 7.15, which shows the code generated for the expectations of each observed m_{vij} value. Constraints were imposed to enable three models to be analyzed using program SURVIV:

model A: ϕ and p constant over years, but p differing between groups;

model B: ϕ constant and the same for both groups; p year-specific and differing between groups;

model C: ϕ year-specific but the same for both groups; p year-specific and differing between groups.

Thus, the unknown parameters of interest are

model A: ϕ
p_v, $v = 1, 2$;

model B: ϕ
p_{vj}, $v = 1, 2$ and $j = 3, ..., 8$;

model C: ϕ_j, $j = 2, ..., 6$
p_{vj}, $v = 1, 2$ and $j = 3, ..., 7$.

In addition, the products $(\phi_7 p_{v8})$ are estimable for model C.

The parameter estimates under model A were $\hat{\phi} = 0.447$ (se = 0.0227), $\hat{p}_1 = 0.252$ (se = 0.0241), and $\hat{p}_2 = 0.813$ (se = 0.0563). Again, these results indicate differing capture probabilities by group. However, this model does not fit the data ($\chi^2 = 63.4$, 19 df, $P < 0.001$). Considerable pooling was required with an associated loss of 20 df.

The survival rate estimate under model B was 0.418 (se = 0.0217). The capture rates for group 1 varied from 0.080 to 0.409 while the range for group 2 was 0.744 to 1.0. The fit of this model was also poor ($\chi^2 = 23.0$, 9 df, $P = 0.006$) in spite of some pooling over cells with small expected values (and, again, a loss of 20 df). Neither model A nor B seems useful in making inference from these data.

Table 7.15. – Listing of the input file to program SURVIV created by program RELEASE (see Part 9 for additional information) for the Lazuli bunting data.

```
INPUT --- proc title 'Al Stokes Bunting data;


INPUT --- proc model npar=24;
INPUT ---   cohort=367 /* Releases for group 1 on occasion 2 */;
INPUT ---     12:s( 1)*s( 7);
INPUT ---     10:s( 1)*(1.-s( 7))*s( 2)*s( 8);
INPUT ---     4:s( 1)*(1.-s( 7))*s( 2)*(1.-s( 8))*s( 3)*s( 9);
INPUT ---     1:s( 1)*(1.-s( 7))*s( 2)*(1.-s( 8))*s( 3)*(1.-s( 9))*s(4)*s(10);
INPUT ---     0:s( 1)*(1.-s( 7))*s( 2)*(1.-s( 8))*s( 3)*(1.-s( 9))*s(4)*(1.-s(10))*s(5)*s(11);
INPUT ---     1:s(1)*(1.-s(7))*s(2)*(1.-s(8))*s(3)*(1.-s(9))*s(4)*(1.-s(10))*s(5)*(1.-s(11))*s(6)*s(12);
INPUT ---   cohort=65 /* Releases for group 1 on occasion 3 */;
INPUT ---     8:s( 2)*s( 8);
INPUT ---     3:s( 2)*(1.-s( 8))*s( 3)*s( 9);
INPUT ---     0:s( 2)*(1.-s( 8))*s( 3)*(1.-s( 9))*s( 4)*s(10);
INPUT ---     0:s( 2)*(1.-s( 8))*s( 3)*(1.-s( 9))*s( 4)*(1.-s(10))*s( 5)*s(11);
INPUT ---     0:s( 2)*(1.-s( 8))*s( 3)*(1.-s( 9))*s( 4)*(1.-s(10))*s( 5)*(1.-s(11))*s( 6)*s(12);
INPUT ---   cohort=230 /* Releases for group 1 on occasion 4 */;
INPUT ---     25:s( 3)*s( 9);
INPUT ---     3:s( 3)*(1.-s( 9))*s( 4)*s(10);
INPUT ---     3:s( 3)*(1.-s( 9))*s( 4)*(1.-s(10))*s( 5)*s(11);
INPUT ---     1:s( 3)*(1.-s( 9))*s( 4)*(1.-s(10))*s( 5)*(1.-s(11))*s( 6)*s(12);
INPUT ---   cohort=255 /* Releases for group 1 on occasion 5 */;
INPUT ---     38:s( 4)*s(10);
INPUT ---     13:s( 4)*(1.-s(10))*s( 5)*s(11);
INPUT ---     1:s( 4)*(1.-s(10))*s( 5)*(1.-s(11))*s( 6)*s(12);
INPUT ---   cohort=256 /* Releases for group 1 on occasion 6 */;
INPUT ---     39:s( 5)*s(11);
INPUT ---     7:s( 5)*(1.-s(11))*s( 6)*s(12);
INPUT ---   cohort=240 /* Releases for group 1 on occasion 7 */;
INPUT ---     46:s( 6)*s(12);
INPUT ---   cohort=31 /* Releases for group 2 on occasion 2 */;
INPUT ---     7:s(13)*s(19);
INPUT ---     4:s(13)*(1.-s(19))*s(14)*s(20);
INPUT ---     0:s(13)*(1.-s(19))*s(14)*(1.-s(20))*s(15)*s(21);
INPUT ---     0:s(13)*(1.-s(19))*s(14)*(1.-s(20))*s(15)*(1.-s(21))*s(16)*s(22);
INPUT ---     0:s(13)*(1.-s(19))*s(14)*(1.-s(20))*s(15)*(1.-s(21))*s(16)*(1.-s(22))*s(17)*s(23);
INPUT ---     0:s(13)*(1.-s(19))*s(14)*(1.-s(20))*s(15)*(1.-s(21))*s(16)*(1.-s(22))*s(17)*(1.-s(23))*s(18)*s(24);
INPUT ---   cohort=23 /* Releases for group 2 on occasion 3 */;
```

Table 7.15 – Continued.

```
INPUT ---    12:s(14)*s(20);
INPUT ---    0:s(14)*(1.-s(20))*s(15)*s(21);
INPUT ---    1:s(14)*(1.-s(20))*s(15)*(1.-s(21))*s(16)*s(22);
INPUT ---    0:s(14)*(1.-s(20))*s(15)*(1.-s(21))*s(16)*(1.-s(22))*s(17)*s(23);
INPUT ---    0:s(14)*(1.-s(20))*s(15)*(1.-s(21))*s(16)*(1.-s(22))*s(17)*(1.-s(23))*s(18)*s(24);
INPUT ---  cohort=34 /* Releases for group 2 on occasion 4 */;
INPUT ---    16:s(15)*s(21);
INPUT ---    3:s(15)*(1.-s(21))*s(16)*s(22);
INPUT ---    0:s(15)*(1.-s(21))*s(16)*(1.-s(22))*s(17)*s(23);
INPUT ---    0:s(15)*(1.-s(21))*s(16)*(1.-s(22))*s(17)*(1.-s(23))*s(18)*s(24);
INPUT ---  cohort=49 /* Releases for group 2 on occasion 5 */;
INPUT ---    20:s(16)*s(22);
INPUT ---    0:s(16)*(1.-s(22))*s(17)*s(23);
INPUT ---    0:s(16)*(1.-s(22))*s(17)*(1.-s(23))*s(18)*s(24);
INPUT ---  cohort=66 /* Releases for group 2 on occasion 6 */;
INPUT ---    28:s(17)*s(23);
INPUT ---    2:s(17)*(1.-s(23))*s(18)*s(24);
INPUT ---  cohort=83 /* Releases for group 2 on occasion 7 */;
INPUT ---    30:s(18)*s(24);
INPUT ---  labels;
INPUT ---    s(1)=Phi(Group=1 Occasion=2);
INPUT ---    s(2)=Phi(Group=1 Occasion=3);
INPUT ---    s(3)=Phi(Group=1 Occasion=4);
INPUT ---    s(4)=Phi(Group=1 Occasion=5);
INPUT ---    s(5)=Phi(Group=1 Occasion=6);
INPUT ---    s(6)=Phi(Group=1 Occasion=7);
INPUT ---    s(7)=p(Group=1 Occasion=3);
INPUT ---    s(8)=p(Group=1 Occasion=4);
INPUT ---    s(9)=p(Group=1 Occasion=5);
INPUT ---    s(10)=p(Group=1 Occasion=6);
INPUT ---    s(11)=p(Group=1 Occasion=7);
INPUT ---    s(12)=p(Group=1 Occasion=8);
INPUT ---    s(13)=Phi(Group=2 Occasion=2);
INPUT ---    s(14)=Phi(Group=2 Occasion=3);
INPUT ---    s(15)=Phi(Group=2 Occasion=4);
INPUT ---    s(16)=Phi(Group=2 Occasion=5);
INPUT ---    s(17)=Phi(Group=2 Occasion=6);
INPUT ---    s(18)=Phi(Group=2 Occasion=7);
INPUT ---    s(19)=p(Group=2 Occasion=3);
INPUT ---    s(20)=p(Group=2 Occasion=4);
```

Table 7.15. – Continued.

```
INPUT --- s(21)=p(Group=2 Occasion=5);
INPUT --- s(22)=p(Group=2 Occasion=6);
INPUT --- s(23)=p(Group=2 Occasion=7);
INPUT --- s(24)=p(Group=2 Occasion=8);
```

The results for model C from program SURVIV are shown in Table 7.16. The average annual survival is estimated to be 0.421 (assumed to be the same for both groups). The goodness of fit statistics are shown in Table 7.17. Several expected values $E(m_{vij})$ are less than two and must be pooled to obtain a test statistic that is more nearly chi-square distributed (e.g., note cohort 8, cell 3 gives $\chi^2 = 14.97$ when one bird was observed while 0.059 was expected). Appropriate pooling results in $\chi^2 = 22.1$, 18 df, $P = 0.228$, indicating a good fit of model C to the data.

The log-likelihood values for models A, B, and C are -95.1064, -66.2558, and -65.2174, respectively. The results of log-likelihood tests between the three models are summarized below:

H_0	H_A	χ^2	df	P
model A	model B	57.70	10	<0.001
model A	model C	59.78	14	<0.001
model B	model C	2.08	4	0.721

These results, taken alone, support the use of either model B or C. The goodness of fit test tends to support only model C.

Ideally, one might want to simulate data similar to those under, at least, models B and C to further understand the performance of the various tests. Alternatively, one could take the parameter estimates $\hat{\phi}_j$ and $\hat{p}_{vj}$ as input into the EXPECT option in program RELEASE. The m_{vij} arrays could be computed and then PROC SURVIV would set up the proper input file for program SURVIV. This would allow bias and precision to be assessed approximately. In addition, the chi-square statistics could be used with Table 3.4 to obtain approximations to the power of tests. The combination of programs RELEASE and SURVIV offer the investigator some powerful analysis tools.

Table 7.16. – Estimates of model parameters for the Lazuli bunting data under model C. Parameters 1-5 and 13-17 are $\phi_j, j = 2, ..., 6$; 7-12 are $\hat{p}_{1j}, j = 3, ..., 7$; and 19-23 are $\hat{p}_{2j}, j = 3, ..., 7$, respectively. Parameters 12 and 24 are $(\phi_7 p_{18})$ and $(\phi_7 p_{28})$, respectively. This output is from program SURVIV.

				95% Confidence Interval	
I	Parameter	S(I)	Standard Error	Lower	Upper
1	1	0.32832895	0.71477979E-01	0.18823211	0.46845279
2	2	0.54410820	0.10040752	0.34730946	0.74090694
3	3	0.41158185	0.66609613E-01	0.28102701	0.54213670
4	4	0.41867222	0.51833132E-01	0.31707928	0.52026516
5	5	0.40113225	0.50680044E-01	0.30179936	0.50046514
6	-23	1.0000000	0.00000000E+00	1.0000000	1.0000000
7	7	0.10198075	0.35142248E-01	0.33101943E-01	0.17085956
8	8	0.19840528	0.53712540E-01	0.93128698E-01	0.30368186
9	9	0.27146268	0.57229006E-01	0.15929383	0.38363154
10	10	0.29674715	0.51903390E-01	0.19501650	0.39847779
11	11	0.40027946	0.64594367E-01	0.27367450	0.52688442
12	12	0.17369512	0.23052430E-01	0.12851235	0.21887788
13	1	0.32832895	0.71477979E-01	0.18823211	0.46842579
14	2	0.54410820	0.10040752	0.34730946	0.74090694
15	3	0.41158185	0.66609613E-01	0.28102701	0.54213670
16	4	0.41867222	0.51833132E-01	0.31707928	0.52026516
17	5	0.40113225	0.50680044E-01	0.30179936	0.50046514
18	-24	1.0000000	0.00000000E+00	1.0000000	1.0000000
19	14	0.53822731	0.18980953	0.16620062	0.91025400
20	15	0.89278880	0.12672562	0.64440658	1.1411710
21	16	0.74418196	0.15468978	0.44098998	1.0473739
22	17	1.0000000	0.36776340E-08	0.99999999	1.0000000
23	6	0.88363930	0.96052303E-01	0.69537679	1.0719018
24	13	0.36914358	0.52853778E-01	0.26555017	0.47273699

Table 7.17. – Goodness of fit statistics for the Lazuli bunting data under model C. This output is from program SURVIV.

Cohort	Cell	Observed	Expected	Chi-square	Note
1	1	12	12.288	0.007	0 < P < 1
1	2	10	11.682	0.242	0 < P < 1
1	3	4	5.273	0.307	0 < P < 1
1	4	1	1.758	0.327	0 < P < 1
1	5	0	0.669	0.669	0 < P < 1
1	6	1	0.174	3.918	0 < P < 1
1	7	339	335.156	0.044	0 < P < 1
2	1	8	7.017	0.138	0 < P < 1
2	2	3	3.168	0.009	0 < P < 1
2	3	0	1.056	1.056	0 < P < 1
2	4	0	0.402	0.402	0 < P < 1
2	5	0	0.105	0.105	0 < P < 1
2	6	54	53.253	0.010	0 < P < 1
3	1	25	25.698	0.019	0 < P < 1
3	2	3	8.568	3.619	0 < P < 1
3	3	3	3.260	0.021	0 < P < 1
3	4	1	0.848	0.027	0 < P < 1
3	5	198	191.625	0.212	0 < P < 1
4	1	38	31.681	1.260	0 < P < 1
4	2	13	12.055	0.074	0 < P < 1
4	3	1	3.137	1.456	0 < P < 1
4	4	203	208.126	0.126	0 < P < 1
5	1	39	41.105	0.108	0 < P < 1
5	2	7	10.697	1.278	0 < P < 1
5	3	210	204.198	0.165	0 < P < 1
6	1	46	41.687	0.446	0 < P < 1
6	2	194	198.313	0.094	0 < P < 1
7	1	7	5.478	0.423	0 < P < 1
7	2	4	2.283	1.291	0 < P < 1
7	3	0	0.084	0.084	0 < P < 1
7	4	0	0.012	0.012	0 < P < 1
7	5	0	0.000	0.000	0 < P < 1
7	6	0	0.000	0.000	0 < P < 1
7	7	20	23.143	0.427	0 < P < 1
8	1	12	11.173	0.061	0 < P < 1
8	2	0	0.411	0.411	0 < P < 1
8	3	1	0.059	14.967	0 < P < 1
8	4	0	0.000	0.000	0 < P < 1

Table 7.17. – Continued.

8	5	0	0.000	0.000	0 < P < 1
8	6	10	11.357	0.162	0 < P < 1
9	1	16	10.414	2.996	0 < P < 1
9	2	3	1.499	1.504	0 < P < 1
9	3	0	0.000	0.000	0 < P < 1
9	4	0	0.000	0.000	0 < P < 1
9	5	15	22.087	2.274	0 < P < 1
10	1	20	20.515	0.013	0 < P < 1
10	2	0	0.000	0.000	0 < P < 1
10	3	0	0.000	0.000	0 < P < 1
10	4	29	28.485	0.009	0 < P < 1
11	1	28	23.394	0.907	0 < P < 1
11	2	2	1.137	0.655	0 < P < 1
11	3	36	41.469	0.721	0 < P < 1
12	1	30	30.639	0.013	0 < P < 1
12	2	53	52.361	0.008	0 < P < 1

7.6 Changes in Group Membership – Desert Tortoise Data

Desert tortoises *Gopherus agassizii* that were uniquely marked were identified near Goffs in eastern San Bernardino County, California, in 1977 and 1980, and between 1983 and 1986 (Turner and Berry, unpublished report, 1986). Only the 1984-1986 data are analyzed in this example.

The purpose of the Goffs study was to estimate sex- and age-specific survival of tortoises for use in the construction of a life table for this species. Carapace lengths and live body masses were recorded for all tortoises registered. The sex of tortoises with carapace lengths less than 180 mm could not be ascertained with certainty, but the sex of all tortoises 180 mm and longer was recorded. Measured carapace lengths ranged from 40 to 325 mm. The size of tortoises affects their susceptibility to capture. Small tortoises are difficult to find and are underrepresented in samples (Berry and Turner 1986). Adult tortoises are conspicuous and have high probabilities of recapture. The size of tortoises also affects survival rates because smaller individuals are more vulnerable to predation by birds, coyotes *Canis latrans*, and kit foxes *Vulpes macrotis*.

Any attempt to estimate survival rates of tortoises should include body size as a variable, and a reasonable approach would be to subdivide the population into groups based on lengths of tortoises measured at time of first capture. Because of possible behavioral and social differences between adult males and females, it would also be desirable to include sex in the

analysis. The Goffs data were divided into 11 groups:

Group number	Sex and size range
1	Males > 208 mm
2	Females > 208 mm
3	Males 180-208 mm
4	Females 180-208 mm
5	155-179 mm
6	140-154 mm
7	120-139 mm
8	100-119 mm
9	80-99 mm
10	60-79 mm
11	< 60 mm

The basic data consist of

$$\begin{array}{lll} R_{v1} & m_{v12} & m_{v13} \\ R_{v2} & & m_{v23} \end{array}$$

for each of the 11 groups (v = 1, ..., 11) plus the average carapace length for each of the groups. The basic data allow estimation of ϕ_{v1}, p_{v2}, and $\phi_{v2}p_{v3}$, giving 33 parameters. We seek a parsimonious model using size to reduce 33 parameters by incorporating growth into the model.

The logistic model provides a reasonable approach to modeling capture and survival probabilities as a function of size. Thus, the survival rate for group 6 during 1984-1985 is expressed as

$$\phi_1 = \frac{\beta_2}{1 + \exp(-\beta_0 - \beta_1 145)},$$

where β_0 is the intercept of the curve, β_1 is the slope and β_2 is the asymptote (≤ 1). The value 145 is the mean carapace length of group 6 tortoises captured in 1984. For group 5, the survival rate for 1984-1985 is

$$\phi_1 = \frac{\beta_2}{1 + \exp(-\beta_0 - \beta_1 171)}$$

because the mean carapace length of group 5 tortoises captured in 1984 was 171 mm. The parameters β_0, β_1, and β_2 are the same as for group 6 and all other groups from 1984 to 1985.

That is, these three parameters are common to all 11 groups because only the mean carapace length for each group is assumed to affect survival. Thus, the estimates of β_0, β_1, and β_2 reflect the environmental conditions of the 1984-85 interval, particularly rainfall. A different set of estimates for β_0, β_1, and β_2 could have been obtained for 1985-1986, and likewise for 1986-1987, had data been collected for all of these periods.

Some tortoises grew 50 mm over a 3-year period. Thus, a mechanism is required in the analysis that allows for the increase in size of tortoises through time – in effect, to allow tortoises to change their group membership. To provide the best estimate of the size of tortoises in 1985 that were in group v in 1984, the mean size of the recaptures in 1985 is calculated, and these values are used in the logistic functions. Further, size in 1985 is used in the logistic function for capture probability in 1985. Thus, for group 5 animals, the probability of recapture in 1985 is

$$p_2 = \frac{\gamma_2}{1 + \exp(-\gamma_0 - \gamma_1 184)}$$

where the mean carapace length of the tortoises classified as group 5 in 1984 is now 184 mm in 1985. Typically, animals in groups 3-11 increased in size over an interval, while the mature adults (groups 1-2) did not.

As with the models described earlier in this monograph, the final pair of parameters ϕ_2 and p_3 are not individually estimable, but the product of the pair can be estimated. Again, this product is treated as a logistic function with three parameters across all 11 groups.

To summarize, the basic model for group v consists of three parameters: ϕ_1, p_2, and the product $\phi_2 p_3$. Each of these three parameters is modeled as a three-parameter logistic function of size. Thus, a total of nine parameters (rather than 33) is estimated from the data, i.e., three logistic functions times three parameters per function.

The basic data for this model are given in Table 7.18. The recaptures are presented as an m_{ij} matrix because this input (Table 7.19) is needed by program SURVIV (White 1983). The second line in each entry is the mean carapace length for tortoises in the cohort. This mean is estimated from all tortoises starting the cohort, even though some of them may be represented in a new cohort because of a previous capture.

Two "tricks" are used to analyze the data with program SURVIV. First, the coding for a logistic function is complex. Rather than code each logistic function separately, a function call, RL, with three arguments is used: the index of the starting parameter for the triplet forming the logistic function, the index of the parameter specifying a correction for adult females (discussed later), and the carapace length used for calculating probability. Instead of separating these values with commas, percent signs (%) are used because SURVIV does not recognize % as a separator. After the compile run of SURVIV is completed, the EST.FOR file SURVIV generates is edited; the percent signs are converted to commas, and the code for the RL function is added. EST.FOR then is compiled and linked into SURVIV to complete the estimation process.

Table 7.18. – The m-array and mean carapace lengths (in parentheses) for desert tortoises captured at Goffs, California, 1984-1986. R_2 includes recaptures from the previous occasion.

Group	R_i	$m_{i,j}$	
		$j=2$	3
1	84	62	7
	(256)	(259)	(261)
	75		51
	(256)		(256)
2	67	57	4
	(222)	(223)	(223)
	71		56
	(222)		(222)
3	10	3	1
	(197)	(198)	(205)
	12		4
	(196)		(196)
4	18	11	4
	(195)	(204)	(204)
	18		12
	(196)		(196)
5	18	12	2
	(171)	(184)	(189)
	13		8
	(168)		(168)
6	11	3	3
	(145)	(161)	(167)
	12		6
	(147)		(147)
7	16	7	1
	(129)	(139)	(153)
	18		6
	(127)		(127)
8	19	10	1
	(107)	(117)	(123)
	27		5
	(109)		(109)
9	16	7	1
	(89)	(98)	(97)
	16		3
	(90)		(90)

Table 7.18. – Continued.

10	9	1	1
	(71)	(80)	(80)
	7		2
	(72)		(72)
11	13	2	1
	(49)	(58)	(64)
	10		0
	(52)		(52)

Table 7.19. – Input to program SURVIV for desert tortoise data.

```
proc title Analysis of tortoise, Section 8 only, sex differences,
84-86 only;
proc model npar=12
 /*rl(i,j,k) converted to
s(i+2)/(1.+exp(-s(i)-s(i+1)*dble(k)-s(j)*dble(k)) */;
 cohort=84 /* Group 1 Males > 208 mm captured 1984 */;
  62:rl(1%0%256)*rl(7%0%259);
   7:rl(1%0%256)*(1.-rl(7%0%259))*rl(4%0%261);
 cohort=67 /* Group 2 Females > 208 mm captured 1984 */;
  59:rl(1%10%222)*rl(7%11%223);
   4:rl(1%10%222)*(1.-rl(7%11%223))*rl(4%12%223);
 cohort=10 /* Group 3 Males 180-208 mm captured 1984 */;
   3:rl(1%0%197)*rl(7%0%198);
   1:rl(1%0%197)*(1.-rl(7%0%198))*rl(4%0%205);
 cohort=18 /* Group 4 Females 180-208 mm captured 1984 */;
  11:rl(1%10%195)*rl(7%11%204);
   4:rl(1%10%195)*(1.-rl(7%11%204))*rl(4%12%204);
 cohort=18 /* Group 5 155-179 mm captured 1984 */;
  12:rl(1%0%171)*rl(7%0%184);
   2:rl(1%0%171)*(1.-rl(7%0%184))*rl(4%0%189);
 cohort=11 /* Group 6 140-154 mm captured 1984 */;
   3:rl(1%0%145)*rl(7%0%161);
   3:rl(1%0%145)*(1.-rl(7%0%161))*rl(4%0%167);
 cohort=16 /* Group 7 120-139 mm captured 1984 */;
   7:rl(1%0%129)*rl(7%0%139);
```

Table 7.19. – Continued.

```
  1:rl(1%0%129)*(1.-rl(7%0%139))*rl(4%0%153);
cohort=19 /* Group 8 100-119 mm captured 1984 */;
 10:rl(1%0%107)*rl(7%0%117);
  1:rl(1%0%107)*(1.-rl(7%0%117))*rl(4%0%123);
cohort=16 /* Group 9 97-99 mm captured 1984 */;
  7:rl(1%0%89)*rl(7%0%98);
  1:rl(1%0%89)*(1.-rl(7%0%98))*rl(4%0%97);
cohort=9 /* Group 10 60-79 mm captured 1984 */;
  1:rl(1%0%71)*rl(7%0%80);
  1:rl(1%0%71)*(1.-rl(7%0%80))*rl(4%0%80);
cohort=13 /* Group 11 < 60 mm captured 1984 */;
  2:rl(1%0%49)*rl(7%0%58);
  1:rl(1%0%49)*(1.-rl(7%0%58))*rl(4%0%64);
cohort=75 /* Group 1 Males > 208 mm captured 1985 */;
 51:rl(4%0%256);
cohort=71 /* Group 2 Females > 208 mm captured 1985 */;
 56:rl(4%12%222);
cohort=12 /* Group 3 Males 180-208 mm captured 1985 */;
  4:rl(4%0%196);
cohort=18 /* Group 4 Females 180-208 mm captured 1985 */;
 12:rl(4%12%196);
cohort=13 /* Group 5 155-179 mm captured 1985 */;
  8:rl(4%0%168);
cohort=12 /* Group 6 140-154 mm captured 1985 */;
  6:rl(4%0%147);
cohort=18 /* Group 7 120-139 mm captured 1985 */;
  6:rl(4%0%127);
cohort=27 /* Group 8 100-119 mm captured 1985 */;
  5:rl(4%0%109);
cohort=16 /* Group 9 97-99 mm captured 1985 */;
  3:rl(4%0%90);
cohort=7 /* Group 10 60-79 mm captured 1985 */;
  2:rl(4%0%72);
cohort=10 /* Group 11 < 60 mm captured 1985 */;
  0:rl(4%0%52);
labels;
 s(1)=Intercept for 1984 survival function;
 s(2)=Slope for 1984 survival function;
 s(3)=Asymptote for 1984 survival function;
 s(4)=Intercept for 1986 survival and cap. prob. function;
 s(5)=Slope for 1986 survival and cap. prob. function;
```

Table 7.19. – Continued.

```
  s(6)=Asymptote for 1986 survival and cap. prob. function;
  s(7)=Intercept for 1985 capture probability function;
  s(8)=Slope for 1985 capture probability function;
  s(9)=Asymptote for 1985 capture probability function;
  s(10)=Adult female survival probability effect;
  s(11)=Adult female capture probability effect;
  s(12)=Adult female 1986 survival and capture probability
effect;
proc estimate novar nsig=5 name=asym_fix;
 initial; s(1)=0.8; s(2)=0.002; s(4)=0.8; s(5)=0.002;
        s(7)=-2.; s(8)=0.02;
 constraints; s(1)>-5.; s(1)<5.; s(2)>0.; s(2)<1.; s(3)=1.0;
  s(4)>-5.; s(4)<5.; s(5)>0.; s(5)<1.; s(6)=1.0;
  s(7)>-5.; s(7)<5.; s(8)>0.; s(8)<1.; s(9)=1.0;
  s(10)=0.0; s(11)=0.0; s(12)=0.0;
proc estimate novar nsig=5 maxfn=1200 name=phi_cons;
 initial; retain=asym_fix;
 constraints; s(1)>-5.; s(1)<5.; s(2)=0.; s(3)=1.0;
  s(4)>-5.; s(4)<5.; s(5)>0.; s(5)<1.; s(6)=1.0;
  s(7)>-5.; s(7)<5.; s(8)>0.; s(8)<1.; s(9)=1.0;
  s(10)=0.0; s(11)=0.0; s(12)=0.0;
proc estimate novar nsig=5 maxfn=1200 name=p_cons;
 initial; retain=asym_fix;
 constraints; s(1)>-5.; s(1)<5.; s(2)>0.; s(2)<1.; s(3)=1.0;
  s(4)>-5.; s(4)<5.; s(5)>0.; s(5)<1.; s(6)=1.0;
  s(7)>-5.; s(7)<5.; s(8)=0.; s(9)=1.0;
  s(10)=0.0; s(11)=0.0; s(12)=0.0;
proc estimate novar nsig=5 maxfn=1200 name=phi&p_cons;
 initial; retain=asym_fix;
 constraints; s(1)>-5.; s(1)<5.; s(2)=0.; s(3)=1.0;
  s(4)>-5.; s(4)<5.; s(5)=0.; s(6)=1.0;
  s(7)>-5.; s(7)<5.; s(8)=0.; s(9)=1.0;
  s(10)=0.0; s(11)=0.0; s(12)=0.0;
proc estimate novar nsig=5 maxfn=1200 name=sex_diff;
 initial; retain=asym_fix;
 constraints; s(1)>-5.; s(1)<5.; s(2)>0.; s(2)<1.; s(3)=1.0;
  s(4)>-5.; s(4)<5.; s(5)>0.; s(5)<1.; s(6)=1.0;
  s(7)>-5.; s(7)<5.; s(8)>0.; s(8)<1.; s(9)=1.0;
proc estimate novar nsig=5 maxfn=5000 name=asymptote;
 initial; retain=asym_fix;
  s(3)=0.95; s(6)=0.95; s(9)=0.95;
```

Table 7.19. – Continued.

```
 constraints;
            s(1)>-5.; s(1)<5.;
            s(4)>-5.; s(4)<5.;
            s(7)>-5.; s(7)<5.;
            s(10)=0.; s(11)=0.; s(12)=0.;
proc estimate novar nsig=5 maxfn=5000 name=s&asymptot;
 initial; retain=asymptote;
 constraints;
            s(1)>-5.; s(1)<5.;
            s(4)>-5.; s(4)<5.;
            s(7)>-5.; s(7)<5.;
proc test;
proc stop;
```

Because of the complexity of the model developed above, a series of models is used to build up to the most complex model. First, estimates are made with program SURVIV for the model labeled ASYM_FIX: the model where the logistic functions for ϕ_1, p_2, and $\phi_2 p_3$ are individually estimated but the asymptotes of each logistic function (β_2, γ_2) are fixed at 1.0. This model only has six parameters; thus it provides a starting point for the estimation process. To test for the effect of growth on survival and capture probabilities, three additional models are included. The model PHI_CONS has $\beta_1 = 0$ for ϕ_1 to test if size affects survival; model P_CONS has $\gamma_1 = 0$ for p_2 to test if size affects capture probability; and PHI&P_CONS has the slope parameter set to zero for all three logistic functions to provide an overall test of the effects of size on survival and capture probabilities.

The final model in the sequence is labeled ASYMPTOTE: each of the three logistic functions in the model are assumed to have different parameter values and the asymptote of each logistic function is estimated rather than fixed at 1.0 (but constrained to <1). Goodness of fit results for this sequence of five models are

Model	Log-likelihood	df	*P*
ASYM_FIX	-70.086	27	0.009
PHI_CONS	-70.478	28	0.020
P_CONS	-75.667	28	<0.001
PHI&P_CONS	-121.125	30	<0.001
ASYMPTOTE	-67.371	24	0.014

None of the five models provided an adequate fit to the observed data as determined by the χ^2 goodness of fit test from program SURVIV. Although size was significantly related to capture probability (P <0.001, likelihood ratio test of P_CONS versus ASYM_FIX), it was not significantly related to survival (P = 0.376, likelihood ratio test of PHI_CONS versus ASYM_FIX). Further, the likelihood ratio test of the general model ASYMPTOTE versus the simpler model ASYM_FIX was not significant (P = 0.143), suggesting that allowing the asymptote values of each of the logistic functions to deviate from 1 did not improve the fit of the model to the observed data. This result is consistent with the poor fit of the ASYMPTOTE model as shown above.

Examination of the partitioned goodness of fit test in the output from program SURVIV showed a pattern in the lack of fit. group 2 (females ≥208 mm) generally contributed large chi-square values, usually with the observed captures exceeding the expected number of captures. This pattern also was visible for group 4 (females 180-208 mm). Further, males in groups 1 and 3 tended to show the opposite pattern – observed captures were generally less than the expected value, although the chi-square contribution from these groups was generally not significant.

These results suggested additional model building to account for sex of the mature tortoises (groups 1-4). Even though males and females might be the same size, behavioral differences might modify their capture and survival probabilities so that size alone would not explain the observed data. Thus, the logistic functions for groups 2 and 4 (females) were modified to include a fourth parameter for survival:

$$\phi_1 = \frac{\beta_2}{1 + \exp\left[-\beta_0 - \beta_1 \text{length} - \beta_3 \max(0, \text{length} - 180)\right]}.$$

The additional parameter β_3 allowed for differential behavior of mature females compared to other groups. Thus, β_3 is zero unless the cohort is in group 2 or 4. This modification added three parameters to the general model described previously because a parameter is needed for each of the three logistic functions.

The three additional parameters caused us to add two additional models to the sequence described above. SEX_DIFF allows for differences in sex, but continues to fix the asymptote value to 1.0 so that the effect of the three additional parameters can be tested against the ASYM_FIX model. The most general model, S&ASYMPTOT has 12 parameters, four for each of the three logistic functions. The goodness of fit results are now much improved:

Model	Log-likelihood	df	P
SEX_DIFF	-58.936	24	0.224
S&ASYMPTOT	-58.219	21	0.201

A summary of the pertinent likelihood ratio tests between models is:

H_0	H_A	χ^2	df	P
PHI_CONS	ASYM_FIX	0.78	1	0.376
P_CONS	ASYM_FIX	11.16	1	<0.001
PHI&P_CONS	ASYM_FIX	102.08	3	<0.001
ASYM_FIX	SEX_DIFF	22.30	3	<0.001
ASYM_FIX	ASYMPTOTE	5.43	3	0.143
ASYM_FIX	S&ASYMPTOT	23.73	6	<0.001
SEX_DIFF	S&ASYMPTOT	1.44	3	0.697
ASYMPTOTE	S&ASYMPTOT	18.30	3	<0.001

We conclude that model SEX_DIFF is the most parsimonious model that fits the observed data. The addition of asymptotes different from unity to the three logistic functions does not improve the fit of the model. The reduced model ASYM_FIX fits the data significantly more poorly than SEX_DIFF, demonstrating that the sex-specific parameters contribute significantly to the fit of the model. Likewise, a test of P_CONS versus ASYM_FIX demonstrates that p is significantly related to the size of the individual. Estimates of the parameters for the SEX_DIFF model are presented in Table 7.20, and plots of the ϕ and p functions are shown in Figure 7.1.

Table 7.20. – Estimates of model parameters for the program SURVIV model SEX_DIFF.

Parameter	Estimate	SE	95% CI
ϕ_1 Intercept	0.476	1.081	-1.643 to 2.595
ϕ_1 Slope	0.00490	0.00487	-0.00465 to 0.0145
ϕ_1 Asymptote	1.0		
p_2 Intercept	-1.405	0.668	-2.714 to 0.0965
p_2 Slope	0.0121	0.00241	0.00739 to 0.0169
p_2 Asymptote	1.0		
$\phi_2 p_3$ Intercept	-2.539	0.250	-3.029 to -2.049
$\phi_2 p_3$ Slope	0.0131	0.00117	0.0108 to 0.0154
$\phi_2 p_3$ Asymptote	1.0		
Adult Females			
ϕ_1 Effect	0.0381	0.0200	-0.00102 to 0.0773
p_2 Effect	0.0207	0.0125	-0.00373 to 0.0452
$\phi_2 p_3$ Effect	0.0238	0.00842	0.00733 to 0.0403

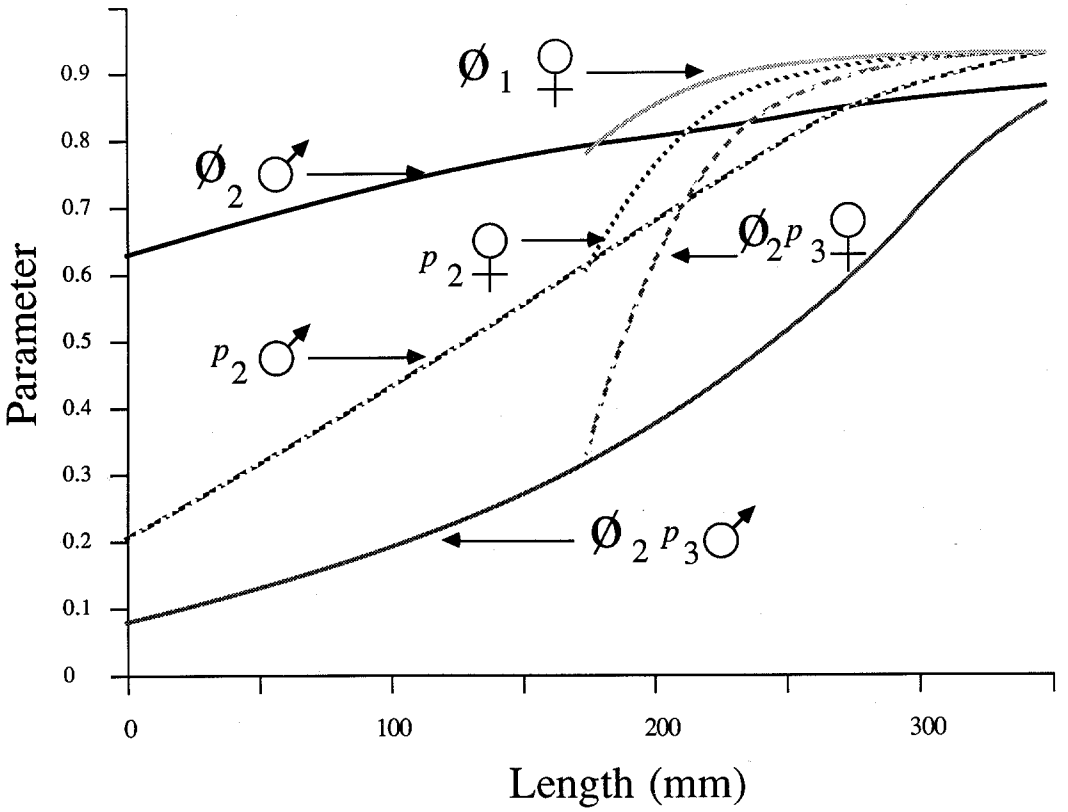

Figure 7.1. – Plots of ϕ_1, p_2, and $\phi_2 p_3$ as a function of carapace length.

Caution must be taken regarding the procedures used in the preceding analysis. The estimates of size for tortoises that were never recaptured may be biased because the probability of capture is related to size. Hence, the tortoises recaptured at a later occasion were probably larger, leading to size-biased sampling. However, because of the large number of groups used in the analysis, the effect of this size-biased effect is minimized. The smaller the size interval used to form a group, the smaller the effect of the size bias. However, adequate sample sizes must be maintained within each group to allow estimation of the group size, and to allow adequate sample sizes for program SURVIV to estimate the multinomial cells.