Part 3. Theory for Studies with Two or More Treatments

3.1. Theory for the Complete Capture History Protocol

As noted in the Reader's Guide (Chapter 1.1), most of Part 3 is intended for persons interested in the theory underlying the methods presented in this monograph. However, biologists cannot safely bypass Chapter 3.10.

3.1.1. Probability Distribution for One Data Set

The starting point for the theory for the complete capture history protocol is the probability distribution for a single data set. That distribution has been considered in the literature (for example, Brownie and Robson 1983; Pollock et al. 1985). Consequently, we do not give the derivations here.

We drop the index for group v in presenting results for a single data set. The complete data are represented as the known releases by subcohorts (subcohorts being determined here by capture histories) and the subsequent recaptures:

$$R_{ih},\ i = 1, ..., k - 1,$$

and

$$m_{ijh},\ j = i + 1, ..., k.$$

Capture history h depends on time of release and ranges over $h = 1, ..., H_i$. The exact set of capture histories being indexed is not relevant to expressing the general theory.

Given R_{ih}, the recaptures $m_{i,i+1,h}, ..., m_{ikh}$ have a multinomial distribution. By assumption, these distributions are independent over subcohorts within i and over cohorts $i = 1, ..., k - 1$. Let the various cell probabilities be π_{ijh}. Also, let $\lambda_{ih} = \sum_{j=i+1}^{k} \pi_{ijh}$ and $r_{ih} = \sum_{j=i+1}^{k} m_{ijh}$ as before. Thus,

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

Then symbolically,

$$\Pr\{\text{Data}\} = \prod_{i=1}^{k-1}\left[\prod_{h=1}^{H_i}\Pr\{m_{i,i+1,h}, \ldots, m_{ikh} \mid R_{ih}\}\right],$$

where

$$\Pr\{m_{i,i+1,h} \cdots m_{ikh} \mid R_{ih}\}$$

$$= \left(\begin{matrix} R_{ih} \\ m_{i,i+1,h} \cdots m_{ikh}\, R_{ih} - r_{ih} \end{matrix}\right)\left[\prod_{j=i+1}^{k}(\pi_{ijh})^{m_{ijh}}\right](1-\lambda_{ih})^{R_{ih}-r_{ih}}.$$

One could consider models wherein the capture probabilities π_{ijh} are dependent on capture history; we do not do so here. Given $\pi_{ij} \equiv \pi_{ijh}$, then the totals $m_{ij} = m_{ij\cdot}$ (and r_i) are a sufficient statistic (minimal sufficient if no particular structure is assumed for the π_{ij}). Thus, it is clear that, given R_i, the m_{ij} are multinomial. For convenience, we refer to the m_{ij} as the cohort recapture data (abbreviated here as cohorts), and write

$$\prod_{i=1}^{k-1}\Pr\{m_{i,i+1}, \ldots, m_{ik} \mid R_i\} = \Pr\{\text{cohorts} \mid \text{releases}\}$$

$$= \prod_{i=1}^{k-1}\left(\begin{matrix} R_i \\ m_{i,i+1} \cdots m_{ik}\, R_i - r_i \end{matrix}\right)\left[\prod_{j=i+1}^{k}(\pi_{ij})^{m_{ij}}\right](1-\lambda_i)^{R_i-r_i}, \quad i = 1, \ldots, k-1.$$

We now can partition the full probability model:

$$\Pr\{\text{Data}\} = \Pr\{\text{subcohorts} \mid \text{cohorts}\}\,\Pr\{\text{cohorts} \mid \text{releases}\}.$$

The conditional distribution of the subcohorts (i.e., all the subcohort data m_{ijh}), given the cohorts, is a series of independent hypergeometric distributions:

Pr{subcohorts|cohorts}

$$= \prod_{i=1}^{k-1} \Pr\{m_{i,i+1,h} \cdots m_{ikh},\ h = 1, ..., H_i \mid m_{i,i+1} \cdots m_{ik}\}$$

$$= \prod_{i=1}^{k-1} \frac{\prod_{h=1}^{H_i} \binom{R_{ih}}{m_{i,i+1,h} \cdots m_{ikh},\ R_{ih} - r_{ih}}}{\binom{R_i}{m_{i,i+1} \cdots m_{ik},\ R_i - r_i}}.$$

We then partition Pr{cohorts|releases} further. Given the classical Jolly-Seber model, which we assume, survival and capture probabilities depend only on the survival interval and recapture occasion, respectively. Hence,

$$\pi_{ij} = \begin{cases} \phi_i p_{i+1}, & j = i + 1 \\ (\phi_i q_{i+1}) \cdots (\phi_{j-2}\, q_{j-1}) \phi_{j-1}\, p_j, & j > i + 1 \end{cases}.$$

The MSS may be taken as $r_1, ..., r_{k-1}, m_2, ..., m_{k-1}$. Its probability distribution is representable as $2k - 3$ conditionally independent binomial distributions:

$$r_i \mid R_i \sim \text{bin}(R_i, \lambda_i), \quad i = 1, ..., k - 1$$
$$m_i \mid T_i \sim \text{bin}(T_i, \tau_i), \quad i = 2, ..., k - 1.$$

For completeness, we reiterate the definitions below:

$$T_2 = r_1;$$

$$T_{i+1} = T_i - m_i + r_i = z_i + r_i, \quad i = 1, ..., k;$$

$$\lambda_i = \phi_i(p_{i+1} + q_{i+1}\lambda_{i+1}), \quad i = 1, ..., k - 1;$$

$$\lambda_k = 0;$$

$$\tau_i = \frac{p_i}{p_i + q_i\lambda_i}, \quad i = 2, ..., k - 1.$$

The conditional distribution of the cohort data given the MSS has been considered by various authors. The earliest derivation appears to be that of Robson and Youngs (unpublished report, 1971). Symbolically we want

$$\Pr\{\text{cohorts}|\text{releases}\} = \Pr\{\text{cohorts}|\text{MSS}\}\ \Pr\{\text{MSS}|\text{releases}\}.$$

Some additional notation is needed. Let m^c_{ij} be the column sum of the m_{lj} for $l = 1, ..., i$:

$$m^c_{ij} = m_{1j} + m_{2j} + \cdots + m_{ij},$$

defined for $i = 1, ..., j - 1$, and $j = 2, ..., k$ (however, only the cases of $i = j - 2$ and $j - 1$ are needed). We note that $z_i = m^c_{i-1,i+1} + ... + m^c_{i-1,k}$, $T_{i+1} = m^c_{i,i+1} + ... + m^c_{ik}$, and $m^c_{i,i+1} = m_{i+1}$. Now

$$\Pr\{\text{cohorts} \mid \text{MSS}\} = \prod_{i=2}^{k-2} \frac{\begin{pmatrix} z_i \\ m^c_{i-1,i+1} \cdots m^c_{i-1,k} \end{pmatrix} \begin{pmatrix} r_i \\ m_{i,i+1} \cdots m_{ik} \end{pmatrix}}{\begin{pmatrix} T_{i+1} \\ m^c_{i,i+1} \cdots m^c_{ik} \end{pmatrix}}.$$

For completeness, the explicit expression for $\Pr\{\text{MSS}|\text{releases}\}$ is

$$\Pr\{\text{MSS} \mid \text{releases}\} = \left[\prod_{i=1}^{k-1} \binom{R_i}{r_i} (\lambda_i)^{r_i} (1 - \lambda_i)^{R_i - r_i}\right] \left[\prod_{i=2}^{k-1} \binom{T_i}{m_i} (\tau_i)^{m_i} (1 - \tau_i)^{T_i - m_i}\right].$$

Thus, we have presented here a partition, for one data set, of $\Pr\{\text{Data}\}$:

$$\begin{aligned}\Pr\{\text{Data}\} &= \Pr\{\text{MSS}|\text{releases}\}\ \Pr\{\text{cohorts}|\text{MSS}\}\ \Pr\{\text{subcohorts}|\text{cohorts}\} \\ &= (\text{component 1}) \times (\text{component 2}) \times (\text{component 3}).\end{aligned}$$

Only component 1 depends on the survival and capture probabilities and is used in deriving the MLEs. Components 2 and 3 are used for goodness of fit tests.

Derivations are easier if one simplifies the notation by defining

$$A_i = \frac{r_i}{R_i},\ i = 1, ..., k - 1,$$

$$B_i = \frac{m_i}{T_i},\ i = 2, ..., k - 1,$$

$A_k = 1$, and $B_k = 1$, and adopting the convention that ϕ_{k-1} means $(\phi_{k-1}p_k)$. The MLEs of λ_1, ..., λ_{k-1} and τ_2, ..., τ_{k-1} are the above A_i and B_i, respectively. The solutions for $\hat{\phi}_i$ and $\hat{p}_i$ are facilitated by writing

$$A_i = \hat{\phi}_i(\hat{p}_{i+1} + \hat{q}_{i+1}A_{i+1}),$$

$$B_{i+1} = \hat{p}_{i+1}/(\hat{p}_{i+1} + \hat{q}_{i+1}A_{i+1}),$$

hence, $A_i = \hat{\phi}_i \hat{p}_{i+1}/B_{i+1}$, and solving for

$$\hat{p}_{i+1} = \frac{B_{i+1}}{B_{i+1} + (1 - B_{i+1})/A_{i+1}}, \quad i = 1, ..., k - 2.$$

Thus,

$$\hat{\phi}_i = A_i(B_{i+1} + (1 - B_{i+1})/A_{i+1}), \quad i = 1, ..., k - 1.$$

The main advantage of these representations of $\hat{\phi}_i$ and $\hat{p}_i$ are that the A_i and B_i are all mutually independent and have known distributions, thereby making it relatively easy to derive variances and covariances of the MLEs. For example, if the delta method is used,

$$\hat{\text{var}}(\hat{\phi}_i) = \left(\frac{\partial\hat{\phi}_i}{\partial A_i}\right)^2 \frac{A_i(1 - A_i)}{R_i} + \left(\frac{\partial\hat{\phi}_i}{\partial B_{i+1}}\right)^2 \frac{B_{i+1}(1 - B_{i+1})}{T_{i+1}}$$
$$+ \left(\frac{\partial\hat{\phi}_i}{\partial A_{i+1}}\right)^2 \frac{A_{i+1}(1 - A_{i+1})}{R_{i+1}}.$$

More detailed consideration of variances and covariances is deferred to Sections 3.1.2 and 3.1.3.

3.1.2. Theory under the Sequence of Models

We outline here the probability theory and inference methods under the sequence of models H_0, $H_{1\phi}$, ..., $H_{k-1,\phi}$ for the complete capture history protocol. We begin with component 1 of the probability distribution Pr{MSS|releases}; and re-introduce the group index v.

Under the most general hypothesis $H_{k-1,\phi}$ of all parameters differing by treatment, the MSS is

$$\text{MSS} = \{\text{MSS}_v, v = 1, ..., V\}$$

or

$$\{r_{v1}, ..., r_{v,k-1}, m_{v2}, ..., m_{v,k-1}, v = 1, ..., V\}.$$

The basic estimable parameters are $\phi_{v1}, ..., \phi_{v,k-2}, (\phi_{v,k-1}p_{vk}), p_{v2}, ..., p_{v,k-1}, v = 1, ..., V$. Each MSS_v is independently and binomially (bin) distributed as in the previous section:

$$r_{vi} \sim \text{bin}(R_{vi}, \lambda_{vi}), \quad i = 1, ..., k-1,$$
$$m_{vi} \sim \text{bin}(T_{vi}, \tau_{vi}), \quad i = 2, ..., k-1.$$

At the other extreme we have H_0: all parameters are the same over treatments. Therefore all $\lambda_{vi} \equiv \lambda_i$ and $\tau_{vi} \equiv \tau_i$. Thus, the MSS under H_0 is simply the sums $r_{\cdot i}$, $i = 1, ..., k-1$ and $m_{\cdot i}$, $i = 2, ..., k-1$, with probability distribution

$$r_{\cdot i} \sim \text{bin}(R_{\cdot i}, \lambda_i), \quad i = 1, ..., k-1,$$
$$m_{\cdot i} \sim \text{bin}(T_{\cdot i}, \tau_i), \quad i = 2, ..., k-1.$$

The theory under Section 3.1.1 applies to obtaining MLEs of the now common parameters.

We can next derive a test of H_0 versus the alternative H_A of $H_{k-1,\phi}$, all parameters may differ by treatments. Let MSS_{H_0} and MSS_{H_A} be the relevant minimal sufficient statistics under null and alternative hypotheses. We want

$$\Pr_{H_0}\{\text{MSS}_{H_A} \mid \text{MSS}_{H_0}\} = \frac{\Pr_{H_0}\left\{\text{MSS}_{H_A}\right\}}{\Pr_{H_0}\left\{\text{MSS}_{H_0}\right\}}.$$

This distribution is a product of $2k - 3$ hypergeometric distributions. For ease of reference to the sequence of hypotheses, we write this distribution in the following order:

$$\left[\frac{\prod_{v=1}^{V}\binom{R_{v1}}{r_{v1}}}{\binom{R_{.1}}{r_{.1}}}\right]\prod_{i=2}^{k-1}\left[\frac{\prod_{v=1}^{V}\binom{T_{vi}}{m_{vi}}}{\binom{T_{.i}}{m_{.i}}}\right]\left[\frac{\prod_{v=1}^{V}\binom{R_{vi}}{r_{vi}}}{\binom{R_{.i}}{r_{.i}}}\right]. \tag{3.1}$$

Let the terms (distributions) here be labeled 1.R1, 1.T2, 1.R2, ..., 1.T_{k-1}, 1.R_{k-1} ($2k$ - 3 of these distributions). These terms produce the corresponding components of TEST 1. In this ordering, terms are aligned with the sequence of alternative hypotheses (models)

$$\begin{array}{ll} H_{1\phi} & (1.R1) \\ H_{2p} & (1.T2) \\ H_{2\phi} & (1.R2) \\ \cdot & \\ \cdot & \\ \cdot & \\ H_{k-1,p} & (1.Tk-1) \\ H_{k-1,\phi} & (1.Rk-1). \end{array}$$

Consideration of the MSS under intermediate hypotheses remains. Although the "bookkeeping" of this process can be confusing, the concept of what occurs is straightforward. The parameters have a natural ordering in time or space. The corresponding sequence of hypotheses produces a series of nested models allowing closed-form tests and estimators. (This sequence of nested models is not unique; this matter is discussed below.) For a given hypothesis such as $H_{2\phi}$, all lower-order parameters are allowed to be different over v: ϕ_{v1}, p_{v2}, and ϕ_{v2}. All higher-order parameters are the same over v: $p_{.3}$, $\phi_{.3}$, ..., $p_{.k-1}$, $(\phi_{.k-1}p_{.k})$, which translates into

$$\lambda_{v1},\ \tau_{v2},\ \lambda_{v2}$$

being different by treatment group but

$$\tau_{.3},\ \lambda_{.3},\ ...,\ \tau_{.k-1},\ \lambda_{.k-1}$$

being the same over v. The above results allow an easy determination of the relevant MSS under $H_{2\phi}$:

$$r_{vi} \sim \text{bin}\,(R_{vi}, \lambda_{vi}), \qquad i = 1, 2,$$

$$m_{v2} \sim \text{bin}\,(T_{v2}, \tau_{v2}), \text{ and}$$

$$r_{.i} \sim \text{bin}\,(R_{.i}, \lambda_{.i}), \qquad i = 3, ..., k-1,$$

$$m_{.i} \sim \text{bin}\,(T_{.i}, \tau_{.i}), \qquad i = 3, ..., k-1.$$

The minimal sufficient statistics, point estimators, and variances and covariances are easily given if one adopts a few conventions. Specifically, let the treatment group index v range over the set $\{1, ..., v, \text{“.”}\}$ where v = “.” denotes pooling over all groups. Thus, in the above expression the MSS is representable as

$$\begin{matrix} r_{v1}, r_{v2}, r_{.3}, ..., r_{.k-1} \\ m_{v2}, m_{.3}, ..., m_{.k-1}, \end{matrix}$$

for $v = 1, ..., V$ or just r_{vi}, $i = 1, ..., k-1$ and m_{vi}, $i = 2, ..., k-1$ with $v = .$ for $i = 3, ..., k-1$. The complete specification of all MSSs, point estimators, and variances and covariances is given in Section 3.1.3.

The probability distribution represented by formula (3.1) is unique; however, the association (interpretation) of intermediate models (between H_0 and $H_{k-1,\phi}$) with terms of this distribution is not unique. A variety of nested models can be created that all give rise to the same sequence of MSSs, and thus to formula (3.1). The biology must dictate the sequence of models one considers. Moreover, there are models (and sequences) that do not produce an MSS of closed form that corresponds to that of any MSS in the sequence $H_0, H_{1\phi}, ..., H_{k-1,\phi}$. In those cases, closed-form results do not exist. Alternative sequences that will lead to closed-form results include $H_{1\phi}, H_{2\phi}, H_{3\phi}, ..., H_{k-1,\phi}$ (i.e., ignore the intermediate cases in H_{ip}) or $H_0, H_{2p}, H_{3p}, ..., H_{k-1,p}$. In the case of $v > 2$, there are subcases within each $H_{i\phi}$ or H_{ip}. Only the extremes have been formulated here: either all groups or no groups were pooled at a given stage. For example, under $H_{1\phi}$ with general V, the parameters of interest are $\phi_{11}, \phi_{21}, ..., \phi_{V1}$. The extremes are $\phi_{v1} = \phi_1$, all v or all ϕ_{v1} are different. However, closed-form results (as for estimators and tests) also exist under any simple subsetting hypothesis such as $\phi_{11} = \phi_{21} = \phi_{31} = \phi_{a1}, \phi_{41} = \cdots = \phi_{v1} = \phi_{b1}$ and $\phi_{a1} \neq \phi_{b1}$. We do not consider such an alternative nor do we consider expanded sequences of nested models in this general discussion of theory.

We now briefly consider the methodology for obtaining estimators and variances and covariances. Under any of these hypotheses (i.e., models), the following statement is true. If we let l = the number of estimable parameters, the MSS has l components representable in the form

$$y_i \sim \text{bin}\,[Y_i, \delta_i(\underline{\theta})], \; i = 1, ..., l.$$

All y_i are mutually independent. The parameters of natural interest are $\underline{\theta}' = (\theta_1, ..., \theta_l)$ and the $\delta_1, ..., \delta_l$ are a one to one transformation of $\theta_1, ..., \theta_l$. The MLEs of the δ_i are

$$\hat{\delta}_i = \hat{\delta}_i(\hat{\underline{\theta}}) = \frac{y_i}{Y_i}, \quad i = 1, ..., l.$$

These l equations can be solved uniquely for the MLE $\hat{\underline{\theta}}$ (see, e.g., Davidson and Solomon 1974). The resulting $\hat{\theta}_i$ are explicit functions of the ratios $a_j = y_j/Y_j$; for example,

$$\hat{\theta}_i = g_i(a_1, ..., a_l), \quad i = 1, ..., l.$$

Note that $E(a_j) = \delta_j$. Asymptotic theoretical variances (if the delta method or the equivalent ML theory approach is used) are

$$\text{var}(\hat{\theta}_i) = \sum_{j=1}^{l} \left(\frac{\partial g_i}{\partial a_j}\right)^2 \frac{\delta_j(1 - \delta_j)}{Y_j}.$$

Covariances are

$$\text{cov}(\hat{\theta}_i, \hat{\theta}_n) = \sum_{j=1}^{l} \left(\frac{\partial g_i}{\partial a_j}\right)\left(\frac{\partial g_n}{\partial a_j}\right) \frac{\delta_j(1 - \delta_j)}{Y_j}.$$

In the above expressions, partial derivatives are evaluated at $E(a_j) = \delta_j$. It is clear that if $\hat{\theta}_i$ and $\hat{\theta}_n$ have no a_j terms in common, then $\text{cov}(\hat{\theta}_i, \hat{\theta}_n) = 0$. In fact, in this case, $\hat{\theta}_i$ and $\hat{\theta}_n$ are independent.

In the capture models of this monograph, the g_i take only two forms. Thus, it is convenient to define the factors

$$G_{ij} = \frac{\partial g_i}{\partial a_j} \sqrt{\frac{\delta_j(1 - \delta_j)}{Y_j}}$$

and get $\text{var}(\hat{\theta}_i) = \sum_{j=1}^{l} (G_{ij})^2$, $\text{cov}(\hat{\theta}_i, \hat{\theta}_n) = \sum_{j=1}^{l} (G_{ij}G_{nj})$. The results in Section 3.1.3 were obtained in this manner.

For the sequence of models considered here, the estimators are all of the forms

$$\hat{\phi}_{vi} = A_{vi}\left[B_{v,i+1} + (1 - B_{v,i+1})/A_{v,i+1}\right], \quad i = 1, ..., k - 1,$$

$$\hat{p}_{vi} = B_{vi}/\left[B_{vi} + (1 - B_{vi})/A_{vi}\right], \quad i = 2, ..., k - 1,$$

$$v \in \{1, ..., V, .\}.$$

Here, $A_{vi} = r_{vi}/R_{vi}$ and $B_{vi} = m_{vi}/T_{vi}$, with $v = .$, meaning $A_{.i} = r_{.i}/R_{.i}$ and $B_{.i} = m_{.i}/T_{.i}$.

More explicit notation is used here for the G-functions, for example

$$G(A_{vi} \mid \hat{\phi}_{vi}) = \left(\frac{\partial \hat{\phi}_{vi}}{\partial A_{vi}}\right)\sqrt{\frac{A_{vi}(1 - A_{vi})}{R_{vi}}}.$$

For the B_{vi}, more than one representation of G is available. These G-functions (evaluated at data values) are given in Table 3.1. To obtain $\hat{\text{var}}(\hat{p}_{vi})$, for example, one has

$$\hat{\text{var}}(\hat{p}_{vi}) = [G(A_{vi} \mid \hat{p}_{vi})]^2 + [G(B_{vi} \mid \hat{p}_{vi})]^2$$

$$= (\hat{p}_{vi}\hat{q}_{vi})^2\left[\frac{1}{r_{vi}} - \frac{1}{R_{vi}} + \frac{1}{m_{vi}} + \frac{1}{z_{vi}}\right], \quad i = 2, ..., k - 1.$$

To obtain theoretical variances and covariances, one substitutes parameters for the estimators and expected values for statistics.

Table 3.1. – Factors for generating variance and covariance formulae for the $\hat{\phi}_{vi}$ and $\hat{p}_{vi}$.

For $\hat{\phi}_{vi}$		For $\hat{p}_{vi}$	
Variable	G(variable $\mid \hat{\phi}_{vi}$)	Variable	G(variable $\mid \hat{p}_{vi}$)
A_{vi}	$\hat{\phi}_{vi}\sqrt{\frac{1}{r_{vi}} - \frac{1}{R_{vi}}}$,	A_{vi}	$\hat{p}_{vi}\hat{q}_{vi}\sqrt{\frac{1}{r_{vi}} - \frac{1}{R_{vi}}}$,
	$i = 1, ..., k-1$		$i = 2, ..., k-1$
$A_{v,i+1}$	$\hat{\phi}_{vi}\hat{q}_{v,i+1}\sqrt{\frac{1}{r_{v,i+1}} - \frac{1}{R_{v,i+1}}}$,	B_{vi}	$\hat{p}_{vi}\hat{q}_{vi}\sqrt{\frac{1}{m_{vi}} + \frac{1}{z_{vi}}}$,
	$i = 1, ..., k-2$		$\equiv \hat{p}_{vi}\hat{q}_{vi}\sqrt{\frac{T_{vi}}{m_{vi}\, z_{vi}}}$,
$B_{v,i+1}$	$\hat{\phi}_{vi}\hat{q}_{v,i+1}(1 - A_{v,i+1})\sqrt{\frac{m_{v,i+1}}{z_{v,i+1}\, T_{v,i+1}}}$,		$i = 1, ..., k-1$
	$i = 1, ..., k-2$		all other $G(. \mid \hat{p}_{vi}) = 0$
	all other $G(. \mid \hat{\phi}_{vi}) = 0$		

The general formula for a variance or covariance of arbitrary parameters θ_1 and θ_2 is

$$\hat{\mathrm{cov}}(\hat{\theta}_1, \hat{\theta}_2) = \sum_{j=1}^{l} G(\text{variable } j \mid \hat{\theta}_1)\, G(\text{variable } j \mid \hat{\theta}_2).$$

The terms in this summation are zero except for variables in common to both parameter estimators. If there are no variables in common, the covariance is zero. As an example, consider $\hat{\mathrm{cov}}(\hat{\phi}_{vi}, \hat{p}_{vi})$. Only A_{vi} is in common here; hence,

$$\hat{\mathrm{cov}}(\hat{\phi}_{vi}, \hat{p}_{vi}) = \hat{\phi}_{vi}\hat{p}_{vi}\hat{q}_{vi}\left(\frac{1}{r_{vi}} - \frac{1}{R_{vi}}\right),\ i = 1, ..., k-1$$

(bear in mind that $\phi_{v,k-1}$ really means $\phi_{v,k-1}\, p_{vk}$).

The final step in the process is to interpret v against the model being considered. The next section presents all point estimators and variances and covariances under all models, given the complete capture history protocol.

3.1.3. Parameter Estimators, Variances, and Covariances

We here succinctly present formulae for all estimable parameters and their associated variances and covariances, under all models, for the complete capture history protocol. Results here apply also to partial capture history protocol scheme A and (with appropriate interpretation) to scheme B and first capture history protocol data. To achieve this generality, we use an abbreviated notation and some conventions. The index for treatment v ranges over the augmented set $\{1, ..., V,.\}$. The case of v = "." represents a special type of "pooling" of statistics and parameters. The chain of models considered remains H_0, $H_{1\phi}$, H_{2p}, ..., $H_{k-1,p}$, and $H_{k-1,\phi}$. The most general model is $H_{k-1,\phi}$, for which one must adopt the convention that the estimator denoted $\hat{\phi}_{v,k-1}$ really estimates the product $(\phi_{v,k-1})(p_{vk})$. Given this convention about $H_{k-1,\phi}$, the key to these simplified representations is (1) separate treatment of the two subsequences $H_{j\phi}$, j = 1, ..., k - 1 and H_{jp}, j = 2, ..., k - 1, and (2) implicit use of the pooling rule to define the MSS and the parameters that vary by group v for any model in the sequence.

3.1.3.1. Pooling rule. – Under the most general model $H_{k-1,\phi}$, the MSS may be taken as r_{vi} given R_{vi}, i = 1, ..., k - 1 and m_{vi} given $T_{vi} = m_{vi} + z_{vi}$, i = 2, ..., k - 1. For all these quantities, v ranges over 1, ..., V. The MSS for any submodel involves some pooling of these statistics over treatment classes v = 1, ..., V. We use the following notation (for any arbitrary i):

$$r_{\cdot i} = \sum_{v=1}^{V} r_{vi} ;$$

$$R_{\cdot i} = \sum_{v=1}^{V} R_{vi} ;$$

$$m_{\cdot i} = \sum_{v=1}^{V} m_{vi} ;$$

$$z_{\cdot i} = \sum_{v=1}^{V} z_{vi} ;$$

$$T_{\cdot i} = m_{\cdot i} + z_{\cdot i} = \sum_{v=1}^{V} (m_{vi} + z_{vi}) = \sum_{v=1}^{V} T_{vi} .$$

Next we define some ratios:

$$A_{vi} = \frac{r_{vi}}{R_{vi}}, \quad i = 1, ..., k-1,$$

$$B_{vi} = \frac{m_{vi}}{T_{vi}}, \quad i = 2, ..., k-1,$$

with $B_{vk} = 1$ and $A_{vk} = 1$ by definition. Pooled versions of these ratios are denoted by $A_{\cdot i}$ and $B_{\cdot i}$ and are defined as

$$A_{\cdot i} = \frac{r_{\cdot i}}{R_{\cdot i}} \quad \text{and} \quad B_{\cdot i} = \frac{m_{\cdot i}}{T_{\cdot i}} .$$

Similar notation regarding pooling applies to parameters, but with a different meaning: ϕ_{vi} and p_{vi} denote that survival, or capture probability, differs over the treatments whereas $\phi_{\cdot i}$ and $p_{\cdot i}$ means these parameters do not differ by treatment. Thus, the notation $\phi_{\cdot i}$ is equivalent to ϕ_i and means $\phi_{1i} = \phi_{2i} = \cdots = \phi_{k-1,i} = \phi_i$. Similarly, $p_{\cdot i}$ means $p_{vi} = p_i$, for all treatments $v = 1, ..., V$.

If these conventions about the parameters are used, the models can be defined in terms of pooling. Under $H_{j\phi}$, the parameters are $\phi_{v1}, ..., \phi_{vj}, \phi_{\cdot j+1}, ..., \phi_{\cdot k-1}$ and $p_{v2}, ..., p_{vj}, p_{\cdot j+1}, ..., p_{\cdot k-1}$. Under H_{jp}, the parameters are $\phi_{v1}, ..., \phi_{v,j-1}, \phi_{\cdot j}, ..., \phi_{\cdot k-1}$ and $p_{v2}, ..., p_{vj}, p_{\cdot j+1}, ..., p_{\cdot k-1}$. Note that H_{1p} is equivalent to H_0. Table 3.2 summarizes this information about model parameters and gives formulae for the number of estimable parameters in each model.

Under the most general model $H_{k-1,\phi}$, there is no pooling of parameters or of the summary statistics. Under the other extreme of H_0, all the parameters and MSS are pooled over v. The pooling rule for the relevant statistics under the sequence of models is given in Table 3.3.

Table 3.2. – Definition of the models in terms of a pooling rule for estimable parameters; also given are formulae for the numbers of estimable parameters by model. Note that here H_0 is equivalent to H_{1p} and that $\phi_{v,k-1}$ means, by convention, the product $(\phi_{v,k-1}p_{vk})$.

Model $H_{j\phi}$, $j = 1, ..., k-1$:

separate by treatment class	number of parameters
$\phi_{v1}, ..., \phi_{vj}$	Vj
$p_{v2}, ..., p_{vj}$	$V(j-1)$
pooled by treatment class	
$\phi_{j+1}, ..., \phi_{k-1}$	$k-1-j$
$p_{j+1}, ..., p_{k-1}$	$k-1-j$

Total estimable parameters = $V(2j-1) + 2(k-j-1)$.

Model H_{jp}, $j = 1, ..., k-1$:

separate by treatment class	number of parameters
$\phi_{v1}, ..., \phi_{v,j-1}$	$V(j-1)$
$p_{v2}, ..., p_{vj}$	$V(j-1)$
pooled by treatment class	
$\phi_{j}, ..., \phi_{k-1}$	$k-j$
$p_{j+1}, ..., p_{k-1}$	$k-j-1$

Total estimable parameters = $2V(j-1) + 2(k-j) - 1$.

Table 3.3. – Definition of the pooling rule for statistics used in parameter estimators, variances, and covariances under the sequence of models H_0 (= H_{1p}), $H_{1\phi}$ to $H_{k-1,\phi}$.

Model $H_{j\phi}$, $j = 1, ..., k-1$:

statistics separate by treatment class:

r_{wi}, R_{wi} and hence A_{wi}, $i = 1, ..., j$

m_{wi}, z_{wi} and hence B_{wi}, $i = 2, ..., j$

statistics pooled over treatment classes:

$r_{\perp}$, $R_{\perp}$ and hence $A_{\perp}$, $i = j+1, ..., k-1$

$m_{\perp}$, $z_{\perp}$ and hence $B_{\perp}$, $i = j+1, ..., k-1$

Model H_{jp}, $j = 1, ..., k-1$:

statistics separate by treatment class:

r_{wi}, R_{wi} and hence A_{wi}, $i = 1, ..., j-1$

m_{wi}, z_{wi} and hence B_{wi}, $i = 2, ..., j$

statistics pooled over treatment classes:

$r_{\perp}$, $R_{\perp}$ and hence $A_{\perp}$, $i = j, ..., k-1$

$m_{\perp}$, $z_{\perp}$ and hence $B_{\perp}$, $i = j+1, ..., k-1$

3.1.3.2. Parameter estimators and their variances. – For every model in the sequence H_0 (= H_{1p}) to $H_{k-1,\phi}$, the parameter estimators have the same form:

$$\hat{\phi}_{vi} = A_{vi}(B_{v,i+1} + (1 - B_{v,i+1})/A_{v,i+1}), \quad i = 1, ..., k - 1,$$

$$\hat{p}_{vi} = \frac{B_{vi}}{B_{vi} + (1 - B_{vi})/A_{vi}}, \qquad i = 2, ..., k - 1,$$

for $v \in \{1, ..., V, .\}$. Under any model, the MLEs are obtained by applying the pooling rules and conventions for that model. One would not want to compute the previous estimators by hand. The value of this representation for the MLEs is the subsequent ease of programming them (as into program RELEASE), and also the investigation of their theoretical properties.

Sampling variances follow:

$$\hat{\text{var}}(\hat{\phi}_{vi}) = (\hat{\phi}_{vi})^2 \left[\left(\frac{1}{r_{vi}} - \frac{1}{R_{vi}} \right) + (\hat{q}_{v,i+1})^2 \left(\frac{1}{r_{v,i+1}} - \frac{1}{R_{v,i+1}} \right) \right.$$

$$\left. + (\hat{p}_{v,i+1}\hat{q}_{v,i+1}) \frac{(1 - A_{v,i+1})^2}{A_{v,i+1}T_{v,i+1}} \right], \qquad i = 1, ..., k - 1, \quad v \in \{1, ..., V, .\};$$

$$\hat{\text{var}}(\hat{p}_{vi}) = (\hat{p}_{vi}\hat{q}_{vi})^2 \left[\frac{1}{r_{vi}} - \frac{1}{R_{vi}} + \frac{1}{m_{vi}} + \frac{1}{z_{vi}} \right], \quad i = 2, ..., k - 1, \quad v \in \{1, ..., V, .\}.$$

If either m_{vi} or z_{vi} is 0 with poor data, then $\hat{p}_{vi} = 0$ or 1 and $\hat{\text{var}}(\hat{p}_{vi}) = 0$.

3.1.3.3. Covariances under model $H_{j\phi}$. – The subsequence of model $H_{j\phi}$ is defined for $j = 1, ..., k - 1$. Most possible covariances are zero. The formulae for the nonzero covariances are fairly simple; however, some are not defined for all values of j. Nonetheless, we give the formulae below in only their most general terms. Their interpretation thus requires that one use the pooling rules of Tables 3.2 and 3.3, and ignore impossible covariances; in all cases $v, v' \in \{1, ..., V, .\}$.

$$\hat{\text{cov}}(\hat{\phi}_{vi}, \hat{\phi}_{v,i+1}) = -\hat{\phi}_{vi}\hat{\phi}_{v,i+1}\hat{q}_{v,i+1} \left(\frac{1}{r_{v,i+1}} - \frac{1}{R_{v,i+1}} \right), \quad i = 1, ..., k - 2.$$

For $j = 1, ..., k - 2$, the number of these covariances is $Vj + k - j - 2$; when $j = k - 1$, the number is $V(k - 2)$. Next,

$$\hat{\text{cov}}(\hat{\phi}_{vj}, \hat{\phi}_{v'j}) = \hat{\phi}_{vj}\hat{\phi}_{v'j}(\hat{q}_{j+1})^2 \left[\left(\frac{1}{r_{\cdot,j+1}} - \frac{1}{R_{\cdot,j+1}}\right) + \left(1 - \frac{r_{\cdot,j+1}}{R_{\cdot,j+1}}\right)^2 \frac{m_{\cdot,j+1}}{z_{\cdot,j+1}T_{\cdot,j+1}}\right],$$

for $j \leq k\text{-}2$ and $v \neq v'$ for $z_{\cdot,j+1} > 0$. If $z_{\cdot,j+1} = 0$ (but $m_{\cdot,j+1} > 0$), clearly $\hat{\text{cov}}(\hat{\phi}_{vj}, \hat{\phi}_{v'j}) = 0$. There are $V(V - 1)2$ of these covariances for $j = 1, ..., k - 2$ (for $j = k - 1$ we have $H_{k-1,\phi}$ and these covariances are all zero).

The nonzero covariances between $\hat{\phi}$ and $\hat{p}$ are:

$$\hat{\text{cov}}(\hat{\phi}_{vi}, \hat{p}_{vi}) = \hat{\phi}_{vi}\hat{p}_{vi}\hat{q}_{vi}\left(\frac{1}{r_{vi}} - \frac{1}{R_{vi}}\right), \quad i = 2, ..., k - 1.$$

There are $V(j - 1) + k - j - 1$ covariances here, for $j = 1, ..., k - 1$. Finally,

$$\hat{\text{cov}}(\hat{\phi}_{vi}, \hat{p}_{v,i+1}) = \hat{\phi}_{vi}\hat{p}_{v,i+1}(\hat{q}_{v,i+1})^2 \left[\left(\frac{1}{r_{v,i+1}} - \frac{1}{R_{v,i+1}}\right)\right.$$

$$\left. + \left(1 - \frac{r_{v,i+1}}{R_{v,i+1}}\right)\left(\frac{1}{z_{v,i+1}}\right)\right], \quad i = 1, ..., k - 2.$$

If $z_{v,i+1} = 0$, then $\hat{\text{cov}}(\hat{\phi}_{vi}, \hat{p}_{v,i+1}) = 0$. The number of covariances here is $Vj + k - j - 2$ for $j = 1, ..., k - 2$ and $V(k - 2)$ for $j = k - 1$.

From the above expression, we find that the total number of nonzero covariances to be computed under model $H_{j\phi}$ is

$$3(Vj + k - j - 2) + \frac{(V - 1)(V - 2)}{2}, \quad \text{for } j = 1, ..., k - 2,$$

and

$$3V(k - 2), \quad \text{for } j = k - 1.$$

3.1.3.4. Covariances under models H_{jp} and H_0. – Formulae for all nonzero covariances under model H_{jp}, $j = 2, ..., k - 1$ follow. Also, the special case of model H_0, where no parameters differ by group index v, is covered by the formal model H_{1p}, i.e., H_{jp} with $j = 1$. Interpretation of these formulae requires that one use the pooling rules of Tables 3.2 and 3.3 and ignore impossible covariances. In all examples, $v,v' \varepsilon \{1, ..., V, .\}$. There are $V(j - 1) + k - j - 1$

of the covariances below:

$$\hat{\text{cov}}(\hat{\phi}_{vi}, \hat{\phi}_{v,i+1}) = -\hat{\phi}_{vi}\hat{\phi}_{v,i+1}\hat{q}_{v,i+1}\left(\frac{1}{r_{v,i+1}} - \frac{1}{R_{v,i+1}}\right), \quad i = 1, ..., k-2.$$

Next,

$$\hat{\text{cov}}(\hat{\phi}_{v,j-1}, \hat{\phi}_{v',j-1}) = (\hat{\phi}_{v,j-1}\hat{\phi}_{v',j-1}\hat{q}_{vj}\hat{q}_{v'j})\left(\frac{1}{r_{\cdot j}} - \frac{1}{R_{\cdot j}}\right), \text{ all } v \neq v'$$

(v or $v' = .$ does not occur here); for $j = 2$ to $k - 1$, there are $V(V-1)2$ of the above covariances. There are no such covariances for $j = 1$.

Under $H_{j\,p}$, there are some nonzero covariances among some p:

$$\hat{\text{cov}}(\hat{p}_{vj}, \hat{p}_{v'j}) = (\hat{p}_{vj}\hat{p}_{v'j}\hat{q}_{vj}\hat{q}_{v'j})\left(\frac{1}{r_{\cdot j}} - \frac{1}{R_{\cdot j}}\right), \text{ all } v \neq v'$$

(but not v or v'). For $j = 2$ to $k - 1$, there are $V(V-1)2$ covariances (they do not exist for $j = 1$).

Next, we have the covariances between $\hat{\phi}$ and $\hat{p}$:

$$\hat{\text{cov}}(\hat{\phi}_{vi}, \hat{p}_{vi}) = (\hat{\phi}_{vi}\hat{p}_{vi}\hat{q}_{vi})\left(\frac{1}{r_{vi}} - \frac{1}{R_{vi}}\right), \quad i = 2, ..., k-1.$$

There are $V(j-1) + k - j - 1$ covariances here for $j = 1, ..., k - 1$. Finally,

$$\hat{\text{cov}}(\hat{\phi}_{vi}, \hat{p}_{v,i+1}) = -\hat{\phi}_{vi}\hat{p}_{v,i+1}(\hat{q}_{v,i+1})^2\left[\left(\frac{1}{r_{v,i+1}} - \frac{1}{R_{v,i+1}}\right)\right.$$

$$\left. + \left(1 - \frac{r_{v,i+1}}{R_{v,i+1}}\right)\left(\frac{1}{z_{v,i+1}}\right)\right], \quad i = 1, ..., k-2;$$

if $z_{v,i+1} = 0$, then $\hat{\text{cov}}(\hat{\phi}_{vi}, \hat{p}_{v,i+1}) = 0$. There are $V(j-1) + k - j - 1$ of these covariances, $j = 1, ..., k - 1$.

Under H_0, a total of $3k - 2$ nonzero covariances can be computed. For H_{jp}, $j = 2, ..., k - 1$, the number of covariances to be computed is $3(V(j - 1) + k - j - 1) + V(V - 1)$.

3.1.4. Goodness of Fit Tests

If (as we assume) the Jolly-Seber model holds separately for each group, then $H_{k-1,\phi}$ or some less general model will fit the data. For each group, the goodness of fit test statistic is the sum of the TEST 2 and TEST 3 statistics. The information in the subcohorts conditional on the cohorts is used in TEST 3 and that contained in the cohorts is used in TEST 2. The overall goodness of fit test statistic is the sum of these statistics for all groups. Hence, it suffices to give the theory for goodness of fit testing for just one group (no "v" index is used).

3.1.4.1. TEST 3. – TEST 3 is based on the probability distribution of the subcohorts given the cohorts. From Section 3.1.1, that distribution is the product of, in general, $k - 1$ multiple hypergeometric probability distributions:

$$\prod_{i=1}^{k-1} \frac{\prod_{h=1}^{H_i} \binom{R_{ih}}{m_{i,i+1,h} \cdots m_{ikh}\, R_{ih} - r_{ih}}}{\binom{R_i}{m_{i,i+1} \cdots m_{ik}\, R_i - r_i}}.$$

In the case when no new (i.e., not previously released) fish enter the study after $i = 1$, then both $H_1 = 1$ and $H_2 = 1$; hence, there are $k - 3$ distributions to consider. We consider only such studies here.

In principle, for occasion i, the corresponding hypergeometric distribution corresponds to a $(k + 1 - i) \times H_i$ contingency table. For this table, one computes a chi-square test of homogeneity, thereby testing H_0: $\pi_{ijh} = \pi_{ij}$, $j = i+1, ..., k$ for all $h = 1, ..., H_i$. However, because the data are usually too sparse to support use of the full chi-square, some pooling is needed to justify the chi-square approximation. A knowledgeable user would have no difficulty in pooling contingency table cells based on the marginals of the table. We built some fixed pooling rules into RELEASE, following essentially the same logic used by Pollock et al. (1985).

The goodness of fit test component based on all subcohorts within cohort i is labeled TEST 3.*Si*. It nominally has $(H_i - 1)(k - i)$ df. As actually computed, the first part of this test is based on partitioning the corresponding distribution into two (multiplicative) components:

$$\frac{\prod_{h=1}^{H_i} \binom{R_{ih}}{r_{ih}}}{\binom{R_i}{r_i}},$$

which leads to TEST 3.*SRi* nominally as a $2 \times H_i$ contingency table; and

$$\frac{\prod_{h=1}^{H_i} \binom{r_{ih}}{m_{i,i+1,h} \cdots m_{ikh}}}{\binom{r_i}{m_{i,i+1} \cdots m_{ik}}},$$

which leads to TEST 3.*Smi* nominally as a $(k - i) \times H_i$ contingency table. The bulk of the data (i.e., information for testing) ends up in the first component, i.e., TEST 3.*SRi*, which tests H_0: $\lambda_{ih} = \lambda_i$, $h = 1, ..., H_i$. Because most of the r_{ih} will be small (note, however, that all $R_{ih} \geq 1$), pooling over capture histories may be needed. When capture probabilities are small, one r_{ih} will dominate the others; specifically, the r_{ih} for individuals released at time 1 and not caught again until occasion i (i.e., $h = \{10 \cdots 01\}$). If that capture history is denoted here as h', the full distribution underlying TEST 3.*SRi* further partitions as

$$\left[\frac{\binom{R_{ih'}}{r_{ih'}}\binom{R_i - R_{ih'}}{r_i - r_{ih'}}}{\binom{R_i}{r_i}}\right]\left[\frac{\prod_{h \neq h'} \binom{R_{ih}}{r_{ih}}}{\binom{R_i - R_{ih'}}{r_i - r_{ih'}}}\right].$$

The first component corresponds to a 2×2 table representing a maximal pooling of the full TEST 3.*SRi* table. Program RELEASE automatically pools down to the 2×2 table and computes TEST 3.*SRi* from it for most data sets. This strategy is usually good. The user can always obtain the full TEST 3.*Si* table and compute the test based on less pooling if that is warranted.

This pooled version of TEST 3.*SRi* tests

$$H_0: \lambda_{ih'} = \left[\sum_{h \neq h'} R_{ih}\lambda_{ih}\right] \Big/ \left[\sum_{h \neq h'} R_{ih}\right].$$

An advantage of a 2 × 2 table is that the corresponding test can be made one-sided. One might test against

$$H_A: \lambda_{ih'} < \left[\sum_{h \neq h'} R_{ih}\lambda_{ih}\right] \Big/ \left[\sum_{h \neq h'} R_{ih}\right].$$

TEST 3.*Smi*, as routinely computed by program RELEASE, is also a predetermined pooling into a 2 × 2 table of what is nominally a $(k - i) \times H_i$ contingency table. The null hypothesis underlying that full table is

$$H_0: \frac{\pi_{ijh}}{\lambda_{ih}} = \frac{\pi_{ij}}{\lambda_i}, \; j = i + 1, ..., k \text{ for all } h = 1, ..., H_i .$$

The collapsed table from which TEST 3.*Smi* is usually computed by RELEASE is derived from the hypergeometric distribution

$$\frac{\binom{r_{ih'}}{m_{i,i+1,h'}}\binom{r_i - r_{ih'}}{r_{ih'} - m_{i,i+1,h'}}}{\binom{r_i}{r_{ih'}}}.$$

The remaining information, if any, that could bear on TEST 3.*Smi* is not used by RELEASE. The actual null hypothesis tested by TEST 3.*Smi* is

$$H_0: \frac{\pi_{i,i+1,h'}}{\lambda_{ih'}} = \left[\sum_{h \neq h'} \frac{r_{ih}\pi_{i,i+1,h}}{\lambda_{ih}}\right] \Big/ \left[\sum_{h \neq h'} r_{ih}\right].$$

3.1.4.2. TEST 2. – There is also goodness of fit information in the cohort data given the Jolly-Seber MSS. The corresponding residual distribution has no unique representation; a convenient form was derived by Robson and Youngs (unpublished report, 1971); (see also Brownie and Robson 1983). TEST 2 is based on $k - 3$ separate contingency tables (hence, $k \geq 4$ is required for TEST 2 to exist for release-recapture data). Each component test, TEST 2.*Ci*, $i = 2, ..., k - 2$, derives from the conditionally independent multiple hypergeometric distribution indexed by i in Pr{cohorts|MSS} =

$$\prod_{i=2}^{k-2} \frac{\begin{pmatrix} z_i \\ m^c_{i-1,i+1} \cdots m^c_{i-1,k} \end{pmatrix} \begin{pmatrix} r_i \\ m_{i,i+1} \cdots m_{ik} \end{pmatrix}}{\begin{pmatrix} T_{i+1} \\ m^c_{i,i+1} \cdots m^c_{ik} \end{pmatrix}} .$$

Thus, TEST 2.*Ci* is computed from a $2 \times (k - i)$ contingency table. Program RELEASE pools as needed, according to the rule of requiring all expected cell values to exceed two. Pooling of the m_{ij} and m^c_{ij} starts from $j = k$ and hence proceeds backwards from the sparser data.

We use π_{ij}, where $E(m_{ij} \mid R_i) = R_i \pi_{ij}$, $j = i + 1, ..., k$, to denote the null hypothesis most easily. Then, for TEST 2.*Ci*, the null hypothesis before any pooling is

$$H_0: \frac{\pi_{ij}}{\sum_{j=i+1}^{k} \pi_{ij}} = \frac{\sum_{n=1}^{i-1} R_n \pi_{nj}}{\sum_{j=i+1}^{k} \sum_{n=1}^{i-1} R_n \pi_{nj}}, \quad j = i + 1, ..., k .$$

(This null hypothesis is true under the Jolly-Seber model.)

As pointed out in Robson and Youngs (unpublished report, 1971), greater power can sometimes be gained by partitioning each TEST 2.*Ci*, especially into a 2×2 table on $j = i + 1$ versus pooling over $j = i + 2, ..., k$. We recommend this partition of TEST 2.*C*.

3.1.4.3. Comment on uniqueness. – Because TEST 2 has been in use for many years, something is known about it. In particular, it has fair to good power against many likely alternatives (such as age effects or behavioral effects to capture). Less is known about the power of TEST 3; however, it seems to be low (see Pollock et al. 1985), especially if the data are sparse. Conceptually, the goodness of fit testing arising from the sum of the chi-square TEST 2 and TEST 3 results is unique. However, this goodness of fit test must be computed as the sum of many components. These components constitute a partition of the overall goodness of fit test. There is no unique way to do this partitioning. There are even alternatives to the major split that we have called TEST 2 and TEST 3 (see, for example, Pollock et al. 1985). In principle, if one knows of a specific alternative hypothesis to Jolly-Seber, a partition of the full goodness of fit test can be found to split out an optimal subcomponent test against that alternative. The situation is analogous to 1 df contrasts in analysis of variance.

3.1.4.4. Comment on optimality. – The tests we present here have some desirable properties. This is true of all of TESTs 1, 2, and 3 despite the lack of uniqueness in their partitioned computational form. These tests are "similar tests" (see Lehmann 1959). This

statistical property is highly desirable; here it just means these tests have the intended α-levels (given the assumptions) regardless of the true unknown parameters. Thus, in testing, for example, $\lambda_{c1} = \lambda_{t1} = \lambda_1$ (which TEST 1.*R*1 does), the significance level of the test is not dependent on the unknown value of λ_1 or on any other unknown parameters if the null hypothesis is true. All these tests also have maximum power. The optimality properties of these tests, under the theoretical models used here, assures us that no better tests can be found.

3.1.5. Tests Between Models (TEST 1)

The test between H_0: no treatment effects (i.e., model H_0) and H_A: model $H_{k\text{-}1,\phi}$ is based on the probability distribution of the MSS under H_A given the MSS under H_0 when H_0 is true. That distribution, given in Section 3.1.2, can be written as the product of $2k$ - 3 hypergeometric distributions:

$$\left[\frac{\prod_{v=1}^{V}\binom{R_{v1}}{r_{v1}}}{\binom{R_{.1}}{r_{.1}}}\right]\prod_{i=2}^{k-1}\left[\frac{\prod_{v=1}^{V}\binom{T_{vi}}{m_{vi}}}{\binom{T_{.i}}{m_{.i}}}\right]\left[\frac{\prod_{v=1}^{V}\binom{R_{vi}}{r_{vi}}}{\binom{R_{.i}}{r_{.i}}}\right].$$

Thus, TEST 1 is conveniently computed as the sum of $2k$ - 3 independent chi-squares, each from a $2 \times V$ contingency table. This representation of TEST 1 is valid without putting any interpretation on each test component. With data from an experiment where the treatment is applied at time 1, the sequence of models we gave in Table 2.2 is reasonable to consider. Each component of TEST 1 then has a clear interpretation in this sequence of models (see Table 2.3). Note, however, that when different alternatives intermediate between models H_0 and $H_{k\text{-}1,\phi}$ are considered, these individual test components may have no interpretive value. Rather, one must then go to numerical methods for model fitting and testing (by way of likelihood ratio tests).

TEST 1.*Ri*, $i = 1, \ldots, k - 1$ is associated with the distribution

$$\frac{\prod_{v=1}^{V}\binom{R_{vi}}{r_{vi}}}{\binom{R_{.i}}{r_{.i}}}.$$

The null hypothesis tested is that model H_{ip} holds (interpret H_{1p} as simply model H_0) versus

the alternative that model $H_{i\phi}$ is true. Expressed in terms of parameters, the actual test is of

$$H_0: \lambda_{vi} = \lambda_i, v = 1, ..., V.$$

With specific null and alternative hypotheses such as these, it is possible that neither hypothesis is true. What happens then is not predictable; it can easily happen that H_0 is not rejected, not because the null model holds, but rather because the alternative model is not appreciably better than the null model.

TEST 1.*Ti*, $i = 2, ..., k - 1$ is associated with the distribution

$$\frac{\prod_{v=1}^{V} \binom{T_{vi}}{m_{vi}}}{\binom{T_{.i}}{m_{.i}}}.$$

The null hypothesis tested is that model $H_{i-1,\phi}$ holds versus the alternative that model H_{ip} is true. Expressed in terms of parameters we actually test

$$H_0: \tau_{vi} = \tau_i, v = 1, ..., V.$$

The powers of these tests can be found, given that one knows the R_{vi} or the T_{vi} and the hypothesized parameters λ_{vi} and τ_{vi}. In practice, the R_{vi}, for $i > 1$, and T_{vi} are not known before a study. (Chapter 3.6 defines a way to handle this situation.) To compute asymptotic power, one first needs the noncentrality parameter, and then either a table of the noncentral chi-square distribution (see Owen 1962) or a way to compute that distribution (PC-SAS and SAS version 5 have the noncentral chi-square distribution as a built-in function; SAS programs are produced by SAS, Incorporated, Cary, North Carolina). For example, the noncentrality parameter for TEST 1.*Ri* (summations on n are over $n = 1, ..., V$) is

$$\sum_{v=1}^{V} \frac{\left(R_{.i}\right)^2 \left(R_{vi}\lambda_{vi} - R_{vi}\frac{(\sum R_{n\,i}\lambda_{n\,i})}{R_{.i}}\right)^2}{R_{vi}(\sum R_{n\,i}\lambda_{n\,i})\,(\sum R_{n\,i}(1-\lambda_{n\,i}))}.$$

The point here is that the power of these tests can be studied analytically; Monte Carlo methods are not required to get information on power, especially at a level of resolution useful for study design (e.g., it suffices to know if power will be large, such as >0.9, versus small, such as <0.5). For study design in fisheries-turbine experiments, the first test that should be examined is TEST 1.*R*1. Computing the power of that test under model $H_{1\phi}$ will give useful

guidance on sample size (similar to results obtained by looking at $se(\hat{S})$ under model $H_{1\phi}$). In the case of only a treatment and control group, any of these tests can be used as one-sided tests, and, correspondingly, one-sided powers can be computed.

3.2. Modifications for Other Protocols

3.2.1. Scheme A, Partial Capture Histories

Essentially no modifications are needed to apply all the theory from the complete capture history protocol to scheme A. Scheme A entails initial batch marks. Upon recapture after first release, a second distinguishing mark is applied which is unique to capture site or time. Thus, the second recaptures (i.e., recaptures after second release) can be distinguished and removed from the study. Under scheme A, most potential capture histories do not exist. However, there are releases R_i at every site, $i = 1, ..., k - 1$. Also, for every R_i, there are recaptures, $m_{i,j}$, at all $j = i + 1, ..., k$. Thus, m-array data exist that are identical in structure to the m-arrays under the complete capture history protocol. The first two cohorts are, in fact, identical to the data under complete capture histories. The recapture counts for cohorts 3, ..., $k - 1$ are slightly smaller than under the complete capture history protocol because releases at occasions $i = 3, ..., k$ all have a single capture history, for example:

Occasion i	h
3	{101}
4	{1001}
5	{10001}

Consequently, all releases R_i at occasion i are from recaptures m_{1i} of fish initially released at occasion 1.

Under the Jolly-Seber assumptions of occasion-specific parameters ($\phi_1, ..., \phi_{k-1}, p_2, ..., p_k$), all the theory for TESTS 1 and 2 applies unchanged. All models in the sequence $H_0, H_{1\phi}, ..., H_{k-1,\phi}$ can be used. Estimation formulae are unchanged. The only modification is that TEST 3 cannot be computed (if new animals were being introduced into the study at each release occasion as per Jolly-Seber capture-recapture, then TEST 3 could be computed).

3.2.2. Scheme B, Partial Capture Histories

Only cohorts 1 and 2 exist under scheme B; however, recaptures exist for all occasions $i = 2, ..., k$. Thus, the m-array is simply

$$\{1\} \quad R_1 \quad m_{12}, m_{13}, ..., m_{1k}$$

$$\{11\} \quad R_2 \; m_{23}, ..., m_{2k}$$

TEST 3 cannot be computed under scheme B. TEST 2 reduces to the single component TEST 2.*C*2. A modified form of TEST 1 can be computed.

There are k estimable parameters, including ϕ_1 and p_2, whereas $\phi_2, ..., \phi_{k-1}, p_3, ..., p_k$ are not separately estimable. It is convenient to take the estimable parameters as $\phi_1, p_2, \lambda_2, \tau_3, ..., \tau_{k-1}$. These are equivalent to $\lambda_1, \lambda_2, \tau_2, \tau_3, ..., \tau_{k-1}$, which is in contrast to the complete capture history or scheme A protocol where one can also estimate $\lambda_3, ..., \lambda_{k-1}$ from r_i / R_i, $i = 3, ..., k - 1$. An MSS is

$$r_1 \mid R_1 \sim \text{bin}(R_1, \lambda_1)$$

$$r_2 \mid R_2 \sim \text{bin}(R_2, \lambda_2)$$

$$m_2 \mid T_2 \sim \text{bin}(T_2, \tau_2)$$

$$m_i \mid T_i \sim \text{bin}(T_i, \tau_i), \quad i = 3, ..., k - 1.$$

The first three components are exactly what one gets by pooling all recaptures, within each cohort, for occasions $j = 3, ..., k$. Thus, if one takes the data as

$$\begin{array}{lll} R_1 & m_{12} & z_2 \\ R_2 & & r_2 \end{array}$$

and sets $k = 3$, all the theory for the complete capture history with $k = 3$ applies. This collapsed representation of the data allows one to get point estimates, variances, tests, etc. on ϕ_1, p_2, and $\phi_2 p_3 = \lambda_2$. (It is not necessary to do this collapsing; it is done here only for its heuristic value in understanding the theory.)

With multiple data sets, one can use models $H_0, H_{1\phi}$, and H_{2p}. The more general models are not useful because the corresponding ϕ_i and p_i are not estimable. If model $H_{2\phi}$ is the true case, the ratio $S_2 = \phi_{t2}/\phi_{c2} = \lambda_{t2}/\lambda_{c2}$ is estimable. However, a complete series of tests does not exist to support strongly the choice of $H_{2\phi}$.

All of the TEST 1.*R*3, ..., 1.*Rk* - 1 components drop out of TEST 1. The remaining components are computed exactly as under the complete capture history protocol. The statistics $m_3, ..., m_{k-1}$ and $T_3, ..., T_{k-1}$ have exactly the same meaning here as under the complete capture history protocol. The exact null hypothesis for TEST 1 components is

TEST	H_0
1.*R*1	$\lambda_{v1} = \lambda_1, v = 1, ..., V$
1.*R*2	$\lambda_{v2} = \lambda_2, v = 1, ..., V$
1.*Ti*	$\tau_{vi} = \tau_i, v = 1, ..., V, i = 3, ..., k - 1$

This hypothesis is the same as for these test components under complete capture histories. The problems that arise here are a result of the absence of the components for TEST 1.*Ri*, $i = 3, ..., k - 1$. Thus, under scheme B, certain deviations from the null hypothesis of TEST 1 cannot be detected. This inability to detect some deviations from the null hypothesis relates to the nonidentifiability of parameters. In principle, one could construct some sets of these parameters (for one or more groups, v), which do not fit Jolly-Seber or which have treatment effects beyond model H_{2p}; yet one cannot detect these cases by testing.

If one views these components of TEST 1 as corresponding to a nested sequence of models, then the alternative to TEST 1.*T*3 is 1.*T*.4, not 1.*R*.3 as under complete capture histories. For 1.*T*4 the alternative is 1.*T*.5 and so forth. Hence, one cannot distinguish models H_{ip} from $H_{i\phi}$. For example, if TEST 1.*T*3 rejected model H_{3p} and all of TESTS 1.*T*4 to 1.*Tk* - 1 did not reject (and goodness of fit was acceptable), one still would not know if the "correct" model was $H_{3\phi}$ or H_{4p} (the matter is then, of course, somewhat academic, as one cannot estimate ϕ_{v3}).

3.2.3. First Capture Histories

The MSS for a single data set is

$$r_1 \mid R_1 \sim \text{bin}\,(R_1, \lambda_1)$$

$$m_j \mid T_j \sim \text{bin}\,(T_j, \tau_j),\ j = 2, ..., k - 1.$$

The above MSS is a subset of the MSS under scheme B, which is itself a subset of the MSS under complete capture histories. The meanings of the m_i and T_i are the same; their exact definitions change as compared with complete capture history: $m_j = m_{1j}, j = 2, ..., k - 1$ and $T_j = m_{1j} + \cdots + m_{1k}$. There is insufficient information to unravel any separate ϕ or p parameters (none of them are estimable). Under $H_{1\phi}$, $S = (\phi_{t1}/\phi_{c1}) = (\lambda_{t1}/\lambda_{c1})$ is estimable. The variance of $\hat{S}$ is the same under model $H_{1\phi}$ for first capture histories as under any of scheme A, B, or complete capture history protocols when $H_{1\phi}$ is true.

No components of TESTs 2 or 3 can be computed. For TEST 1, components 1.*R*1 and 1.*Ti*, $i = 2, ..., k - 1$ exist. TEST 1.*R*1 tests the null hypothesis that $\lambda_{v1} = \lambda_1$ for all $v = 1, ..., V$. TEST 1.*Ti* tests the null hypothesis that $\tau_{vi} = \tau_i$ for all $v = 1, ..., V$. As under scheme B, these

tests cannot detect certain alternatives because of the nonidentifiability of the individual parameters. Under the first capture history protocol, one hopes that model $H_{1\phi}$ holds, in which case TEST 1.*R*1 should reject H_0 and all of TESTS 1.*T*2, ..., 1.*Tk* - 1 should not reject. In a general sense the sum of the test statistics for these 1.*Ti* series is a goodness of fit test to the overall model $H_{1\phi}$.

The identifiability problem is eliminated when the capture probabilities are not affected by the treatment; then $p_{1i} = p_{2i} = ... = p_{vi}$ for all $i = 1, ..., k$. The sequence of models under the first (or unknown) capture history protocol then simplifies to $H_{1\phi}, H'_{2\phi}, ..., H'_{k-1,\phi}$. The parameters of model $H'_{i\phi}$ for $i = 2$ to k - 1, are

$$\phi_{v1}, ..., \phi_{vi}, \quad v = 1, ..., V,$$

$$\phi_{i+1}, ..., \phi_{k-1},$$

and

$$p_2, ..., p_k .$$

Given this restrictive *assumption* about the capture probabilities, the sum of the chi-squares of TEST 1.*T*i through TEST 1.*Tk* - 1 provides a test of the null hypothesis that $\phi_{vj} = \phi_j$ for $v = 1$, ..., V and j = i, ..., k - 1.

3.2.4. Some Extensions

3.2.4.1. Relationship to temporal banding studies. – The theory for the first capture history protocol can be directly applied to certain types of experiments based on banded birds. In particular, the sequence of models described as $H_{1\phi}, H'_{2\phi}, H'_{3\phi}, ..., H'_{k-1,\phi}$ are applicable. Brownie et al. (1985) gave extensive background on the analysis of banding data. There are very close links and similarities between band recovery and recapture theory. Both methods are studying survival processes. It is mainly the resampling process that differs between the two types of studies. This difference translates into a different parameterization for expected values of the m_{vij}, which are either recaptures or band recoveries.

Let a banding experiment involve preseason release of treatment and control groups ($V = 2$). Treatment might be forced ingestion of lead pellets (see, for example, Deuel 1985). Then m_{vij} are the band recoveries in year j after banding. There are recoveries in year 1 (direct recoveries), as well as in years $j = 2, ..., k$. The model structure, the $E(m_{vij})/R_{v1}$, for treatment and controls is

	Proportion of bands recovered in year j				
Banded	1	2	3	...	k
R_{t1}	f_{t1}	$\phi_{t1}f_{t2}$	$\phi_{t1}\phi_{t2}f_{t3}$	...	$\phi_{t1}\cdots\phi_{t,k-1}f_{tk}$
R_{c1}	f_{c1}	$\phi_{c1}f_{c2}$	$\phi_{c1}\phi_{c2}f_{c3}$	...	$\phi_{c1}\cdots\phi_{c,k-1}f_{ck}$

The f_{vj} is a recovery rate; ϕ_{vj} is survival rate from year j to $j + 1$.

Treatment might affect the direct recovery rates f_{v1}. This can be tested with a 2×2 table:

m_{t1}	$R_{t1} - m_{t1}$
m_{c1}	$R_{c1} - m_{c1}$

Then, dropping the first year (direct) recoveries, one has data analogous to recapture data under the first capture history protocol. In particular, assuming no treatment effect on recovery rates for $j < 1$ (hence, $f_{vj} = f_j$) and an acute effect, so ϕ_{t1} and ϕ_{c1} differ, but $\phi_{tj} = \phi_{cj} = \phi_j$ for $j > 2$, gives exactly the general structure of model $H_{1\phi}$ for first capture history recapture data. In particular, we can then define the treatment effect as $S = \phi_{t1}/\phi_{c1}$ and we have

$$\frac{E(m_{tj})}{R_{t1}} = S\,\frac{E(m_{cj})}{R_{c1}}, \quad j = 2, ..., k.$$

The m_{vj} are multinomial random variables and the two released groups (cohorts) are independent. Therefore, all the theory for the first capture history model $H_{1\phi}$ is directly applicable. Moreover, that theory extends to goodness of fit testing and exploring the sequence of models $H_{1\phi}, H'_{2\phi}, ..., H'_{k-1,\phi}$.

In banding studies, long time periods are involved; recoveries accrue over years and may be obtained from a spatially unrestricted area. In fisheries studies regarding the effect of a turbine or bypass, the temporal component is limited, and recaptures accrue at known spatial points. However, in both cases we are dealing with resampling cohorts of marked animals exposed to a survival process with possibly a treatment structure imposed on the released cohorts. A common general statistical theory underlies such release-resampling studies of survival processes.

3.2.4.2. Deeper insights. – A unified theory can be given for capture-recapture, release-resampling, bird-banding, and some related types of studies. Such a theory is given by Burnham (unpublished report, 1987); we give here a central feature of this unification (see also Brownie et al. 1985).

Let R marked animals be released as a cohort, at time 1, and then resampled at or after specific subsequent times. The resample counts are $m_2, m_3, ..., m_k$. We require only that the animals survive until time j, $j = 2, ..., k$, to be counted. Then, for some time interval at, or after, j and before $j + 1$, a sampling process occurs to make counts of survivors at time j. Let β_j = the probability of surviving from time j to $j + 1$, given the animal is alive at time j. Let α_j = the probability of being sampled (counted) in the jth sampling interval given the animal was alive at time j. Then the general model structure is

$$E(m_j) = \begin{cases} R\alpha_1, & j = 1 \\ R\beta_1 \cdots \beta_{j-1}\alpha_j, & j > 1 \end{cases}.$$

This is also the structure of banding data (with $\alpha_j = f_j$ and $\beta_j = \phi_j$). Band recoveries may occur continuously during a large part of time interval j to $j + 1$.

The model structure for release-recapture with $k + 1$ occasions is

$$E(m_j) = \begin{cases} R(\phi_1 p_2), & j = 2 \\ R(\phi_1 q_2) \cdots (\phi_{j-2} q_{j-1})(\phi_{j-1} p_j), & j > 2 \end{cases}.$$

A standardized structure occurs if we define

$$\alpha_{j-1} = \phi_{j-1} p_j, \quad j = 2, ..., k + 1, \text{ and}$$

$$\beta_{j-1} = \phi_{j-1} q_j, \quad j = 2, ..., k + 1.$$

With this definition, we have also shifted the indexing so that now we can write

$$E(m_{j+1}) = \begin{cases} R\alpha_1, & j = 1 \\ R\beta_1 \cdots \beta_{j-1}\alpha_j, & j = 2, ..., k \end{cases}.$$

Thus, with a shift of indexing, the standardized model structure is the same for banding data as release-recapture data. Moreover, in either case, the cohorts are multinomial data. Consequently, all underlying theory based on the standardized parameters and indexing is identical for capture-recapture and banding data under the assumptions of time-specific parameters. More general assumptions about parameters are possible and the equivalance of the two processes still holds.

For capture-recapture, β_j is survival in the released cohort; ϕ_{j-1} represents physical survival, and q_j represents not being captured. Once an animal is captured, it is removed from that cohort (of its last release) at risk of capture. Conversely, $\phi_{j-1}p_j$ represents a sampling rate conditional on being alive in the release cohort at time j - 1. Typically in capture-recapture, the resampling occurs at the end of the period j - 1 to j, rather than spread out over the interval, as in band recovery.

As far as abstract statistical theory is concerned, the only difference between capture-recapture, as regards survival estimation, and bird-banding is the interpretation of the standardized parameter presentations:

Standardized parameter	Banding	Capture-recapture
cohort sampling rate	$\alpha_j = f_j$	$\alpha_{j-1} = \phi_{j-1}p_j$
cohort survival rate	$\beta_{j-1} = \phi_j$	$\beta_{j-1} = \phi_{j-1}q_j$
	$j = 1, ..., k$	$j = 2, ..., k + 1$

(the range k is arbitrary, so shifting the indexing is trivial).

One can take capture-recapture data, as an m-array, analyze it with program ESTIMATE, and get Jolly-Seber MLEs as, for example,

$$\hat{\phi}_1 = \hat{\alpha}_1 + \hat{\beta}_1$$

(in Brownie et al. 1985 and ESTIMATE, the notation used is $\hat{f}_1$ for $\hat{\alpha}_1$ and $\hat{S}_1$ for $\hat{\beta}_1$).

3.3. Variances and Covariances of $\hat{S}$

3.3.1. Some Variance Formulae

If a general treatment effect is defined as

$$\hat{S}_i(v, v') = \frac{\hat{\phi}_{vi}}{\hat{\phi}_{v'i}}, \; i = 1, ..., k - 2, \; v \neq v' ,$$

the theoretical asymptotic variance of $\hat{S}_i$ is

$$\text{var}(\hat{S}_i) = (S_i)^2 \left\{ [\text{cv}(\hat{\phi}_{vi})]^2 + [\text{cv}(\hat{\phi}_{v'i})]^2 - 2 \frac{\text{cov}(\hat{\phi}_{vi}, \hat{\phi}_{v'i})}{\phi_{vi}\, \phi_{v'i}} \right\}.$$

Of special interest to us is the case of models $H_{1\phi}$, H_{2p}, and $H_{2\phi}$ for complete capture histories or partial capture history scheme A (the formula below can also apply for partial capture history scheme B). For simplicity, we use t and c rather than v and v' and give $\text{cv}(\hat{S})$ rather than $\text{var}(\hat{S})$. Under model $H_{1\phi}$:

$$[\text{cv}_1(\hat{S})]^2 = \frac{1}{E(r_{t1})} - \frac{1}{R_{t1}} + \frac{1}{E(r_{c1})} - \frac{1}{R_{c1}}.$$

Under model H_{2p}:

$$[\text{cv}_2(\hat{S})]^2 = [\text{cv}_1(\hat{S})]^2 + (p_{t2} - p_{c2})^2 \left[\frac{1}{E(r_{\cdot 2})} - \frac{1}{E(R_{\cdot 2})} \right]$$

$$+ \frac{(1 - \lambda_2)^2}{\lambda_2} \left[\frac{p_{t2}\, q_{t2}}{E(T_{t2})} + \frac{p_{c2}\, q_{c2}}{E(T_{c2})} \right].$$

Under model $H_{2\phi}$:

$$[\text{cv}_3(\hat{S})]^2 = [\text{cv}_1(\hat{S})]^2 + \sum_{v=c}^{t} \left[\frac{p_{v2} q_{v2} (1 - \lambda_{v2})^2}{\lambda_{v2}\, E(T_{v2})} \right]$$

$$+ \sum_{v=c}^{t} (q_{v2})^2 \left[\frac{1}{E(r_{v2})} - \frac{1}{E(R_{v2})} \right].$$

An interesting way to use these specific results is to look at efficiency. Say, for example, that model $H_{1\phi}$ is true but either H_{2p} or $H_{2\phi}$ is considered to gain greater robustness (after all, it is not known that $H_{1\phi}$ is true with real data). Evaluation of these coefficients of variation when $H_{1\phi}$ is true yields $[\text{cv}_1(\hat{S})]^2$ as above, but for models H_{2p} and $H_{2\phi}$ (cv_2 and cv_3, respectively):

$$[\mathrm{cv}_2(\hat{S})]^2 = [\mathrm{cv}_1(\hat{S})]^2 + \frac{(1-\lambda_2)^2 p_2 q_2}{\lambda_2}\left[\frac{1}{E(r_{t1})} + \frac{1}{E(r_{c1})}\right],$$

$$[\mathrm{cv}_3(\hat{S})]^2 = [\mathrm{cv}_2(\hat{S})]^2 + (q_2)^2\left[\frac{1}{E(r_{t2})} - \frac{1}{E(R_{t2})} + \frac{1}{E(r_{c2})} - \frac{1}{E(R_{c2})}\right].$$

From these formulae it becomes clear that the efficiency (regarding $\hat{S}$) when model H_{2p} rather than $H_{1\phi}$ is used (if $H_{1\phi}$ is true) is high, 80 to 90%. However, there is a large loss of efficiency in going to model $H_{2\phi}$; that efficiency, relative to model $H_{1\phi}$, is about 20 to 30%.

3.3.2. Covariances

In this section we index treatments as v, $v1$, $v2$, etc., and thus,

$$S_i(v, v1) = \frac{\phi_{vi}}{\phi_{v1i}}.$$

Of special interest to us are covariances under model $H_{1\phi}$ ($V > 2$) wherein

$$\hat{S}_1(v, v1) = \frac{r_{v,1} / R_{v,1}}{r_{v1,1} / R_{v1,1}};$$

$$\hat{S}_1(v2, v3) = \frac{r_{v2,1} / R_{v2,1}}{r_{v3,1} / R_{v3,1}}.$$

If it is assumed that all v, $v1$, $v2$, and $v3$ are distinct treatments, then $\mathrm{cov}[\hat{S}_1(v, v1), \hat{S}_1(v2, v3)] = 0$. The only nontrivial cases are when the two different $\hat{S}_1$ depend on only three treatment levels, i.e., v, $v1$, and $v2$. If v is made the treatment index in common, e.g., $\hat{S}_1(v, v1)$ and $\hat{S}_1(v, v2)$, only two different formulae arise. In case 1, v is in either both denominators or both numerators and the covariance is

$$\mathrm{cov}[\hat{S}_1(v, v1), \hat{S}_1(v, v2)] = S_a S_b\left[\frac{1}{E(r_{v,1})} - \frac{1}{R_{v,1}}\right];$$

where S_a and S_b represent the two different treatment effects. In case 2, v is in the numerator of one S_a and the denominator of the other S_b and the covariance is

$$\text{cov}[\hat{S}_1(v, v1), \hat{S}(v2, v)] = -S_a S_b \left[\frac{1}{E(r_{v,1})} - \frac{1}{R_{v,1}}\right].$$

In the completely general case,

$$\text{cov}[\hat{S}_i(v, v1), \hat{S}_j(v2, v3)] = (S_i S_j)\left[\frac{\text{cov}(\hat{\phi}_{vi}, \hat{\phi}_{v2j})}{\phi_{vi}\phi_{v2j}} - \frac{\text{cov}(\hat{\phi}_{vi}, \hat{\phi}_{v3j})}{\phi_{vi}\phi_{v3j}}\right.$$

$$\left. - \frac{\text{cov}(\hat{\phi}_{v1i}, \hat{\phi}_{v2j})}{\phi_{v1i}\phi_{v2j}} + \frac{\text{cov}(\hat{\phi}_{v1i}, \hat{\phi}_{v3j})}{\phi_{v1i}\phi_{v3j}}\right].$$

This covariance formula is needed to get variances of products. For example, one might have the product $\hat{S}_1\hat{S}_2$ as the estimate of the treatment effect, where

$$\hat{S}_1 = \hat{\phi}_{t1}/\hat{\phi}_{c1}\,; \quad \hat{S}_2 = \hat{\phi}_{t2}/\hat{\phi}_{c2}\,.$$

Then

$$\text{var}(\hat{S}_1\hat{S}_2) = (S_1S_2)^2\left\{[\text{cv}(\hat{S}_1)]^2 + [\text{cv}(\hat{S}_2)]^2 + 2\,\frac{\text{cov}(\hat{S}_1, \hat{S}_2)}{S_1S_2}\right\}.$$

This asymptotic variance formula can be extended to general products. However, such approximate variance (and covariance) formulae, and the assumption of approximate normality, are usually poor when applied to extended products such as $\hat{S} = \hat{S}_1 \cdots \hat{S}_j = (\hat{\phi}_{t1} \cdots \hat{\phi}_{tj})/(\hat{\phi}_{c1} \cdots \hat{\phi}_{cj})$. A superior procedure is to make inferences based on the log-transformation, $\ln(\hat{S})$. Section 3.5.1 gives general variance and covariance formulae related to $\ln(\hat{S})$.

3.4. Adjustments for Statistical Bias

Parameter estimators have some statistical bias even when the model used as a basis for analysis is true (Gilbert 1973). The survival rate estimator $\hat{\phi}$ is easily adjusted to be unbiased (assuming the model is true). The estimators of p and S can also be easily modified to reduce statistical bias.

There is an option in program RELEASE to print these bias-adjusted estimators. We do not dwell on statistical bias or its adjustment, however, because statistical bias is a trivial source of bias. The serious source of bias is "model" bias. By model biases, we mean biases which occur because the incorrect model is used. Statistical biases are smaller than one standard error of the parameter estimator; however, model biases can be large and thus serious if

the wrong model is used. Our emphasis on model selection and tests of assumptions is designed to minimize model bias. Other significant bias sources are failures of assumptions resulting from tag loss, errors in data recording, and errors in knowing the exact numbers of animals released.

3.4.1. Survival Rate Estimators, $\hat{\phi}$

In all models, $\hat{\phi}$ is of the form

$$\hat{\phi}_{vi} = A_{vi}\left[B_{v,i+1} + (1 - B_{v,i+1})/A_{v,i+1}\right], \quad i = 1, ..., k - 1,$$

where

$$A_{vi} = r_{vi}/R_{vi} \text{ and } B_{vi} = m_{vi}/T_{vi}, \text{ and } v \in \{1, ..., V, .\}.$$

The bias-adjusted estimator of ϕ_{vi} is

$$\tilde{\phi}_{vi} = A_{vi}\left[B_{v,i+1} + (1 - B_{v,i+1})\Big/\left(\frac{r_{v,i+1} + 1}{R_{v,i+1} + 1}\right)\right].$$

The large sample variance of $\tilde{\phi}_{vi}$ is the same as $\hat{\phi}_{vi}$ (the MLE). The expected value of $\tilde{\phi}_{vi}$ is exactly (assuming the correct model)

$$E(\tilde{\phi}_{vi}) = \phi_{vi}\left[1 - q_{v,i+1}\,(1 - \lambda_{v,i+1})^{R_{v,i+1}+1}\right].$$

Usually $\lambda_{v,i+1}$ and $R_{v,i+1}$ are jointly large enough to ensure that the bias in $\tilde{\phi}_{vi}$ is virtually zero (e.g., $\lambda_{v,i+1} \geq 0.05$ and $R_{v,i+1} \geq 99$ suffices).

3.4.2. Capture Probability Estimators, $\hat{p}$

The MLE of p_{vi} is always of the form

$$\hat{p}_{vi} = \frac{B_i}{B_i + (1 - B_i)/A_i}, \; i = 2, \ldots, k - 1.$$

The method that works to adjust the $\hat{\phi}_{vi}$ does not work for $\hat{p}_{vi}$.

Consider

$$\frac{1}{\hat{p}_{vi}} = 1 + \frac{1}{A_i}\left(\frac{1}{B_i} - 1\right).$$

Because A_i and B_i are independent, a good bias adjustment for $\frac{1}{\hat{p}_{vi}}$ is

$$\frac{1}{\tilde{p}_{vi}} = 1 + \left(\frac{R_{vi} + 1}{r_{vi} + 1}\right)\left(\frac{T_{vi} + 1}{m_{vi} + 1} - 1\right),$$

$$E\left(\frac{1}{\tilde{p}_{vi}}\right) = 1 + \frac{1}{\lambda_{vi}}\left[1 - (1 - \lambda_{vi})^{R_{vi}+1}\right]\frac{1}{\tau_{vi}}\left\{\left[1 - (1 - \tau_{vi})^{T_{vi}+1}\right] - 1\right\}.$$

Most situations will justify the following as a good approximation

$$E\left(\frac{1}{\tilde{p}_{vi}}\right) \doteq 1 + \frac{1}{\lambda_{vi}}\left(\frac{1}{\tau_{vi}} - 1\right) = \frac{1}{p_{vi}}.$$

A Taylor's series expansion now leads to

$$E(\tilde{p}_i) \doteq p_{vi}\,[1 + (\mathrm{cv}(\tilde{p}_i))^2\,].$$

Thus, to a first order of approximation

$$E(\tilde{p}_{vi}) \doteq p_{vi}\,[1 + (\mathrm{cv}(\hat{p}_i))^2\,],$$

and a bias-adjusted estimator of p_{vi} is

$$\tilde{\tilde{p}}_{vi} = \frac{\tilde{p}_{vi}}{1 + [\mathrm{cv}(\hat{p}_i)]^2}.$$

3.4.3. Treatment Effect Estimators, $\hat{S}$

The general form of $\hat{S}_i$ is

$$\hat{S}_i = \frac{\hat{\phi}_{ti}}{\hat{\phi}_{ci}}, \quad i = 1, \ldots, k-2.$$

Here, t and c can be interpreted as any v, v'. One might consider using $\tilde{\phi}_{ti} / \tilde{\phi}_{ci}$ as a bias-adjusted estimator of S_i; however, this does not work. For example, under model $H_{1\phi}$, the complete capture history case, we have

$$\hat{S}_1 = \frac{r_{t1} / R_{t1}}{r_{c1} / R_{c1}} = \frac{\hat{\phi}_{t1}}{\hat{\phi}_{c1}} \equiv \frac{\tilde{\phi}_{t1}}{\tilde{\phi}_{c1}};$$

hence, bias-correcting the $\hat{\phi}_{v1}$ need have no effect on the bias of $\hat{S}_1$.

In the previous example, a bias-adjusted $\hat{S}_1 \equiv \hat{S}$ is

$$\tilde{S} = \frac{r_{t1}}{R_{t1}} \Big/ \frac{r_{c1} + 1}{R_{c1} + 1}.$$

For estimators of S under model $H_{1\phi}$, $\tilde{S}$ above is the appropriate bias-adjusted estimator.

For a completely general adjustment, we use a Taylor's series expansion to get the first order approximation

$$E(\hat{S}_i) \doteq S_i\left\{1 + [\mathrm{cv}(\hat{\phi}_{ci})]^2 - S_i\,\frac{\mathrm{cov}(\hat{\phi}_{ti}, \hat{\phi}_{ci})}{\phi_{ti}\,\phi_{ci}}\right\}.$$

Consequently, a generally valid bias-adjusted estimator of S_i is

$$\tilde{S}_i = \frac{\hat{S}_i}{1 + [\text{cv}(\hat{\phi}_{ci})]^2 - \hat{S}_i \dfrac{\hat{\text{cov}}(\hat{\phi}_{ti}, \hat{\phi}_{ci})}{\hat{\phi}_{ti}\,\hat{\phi}_{ci}}}.$$

3.5. Transformations of $\hat{S}$, $\hat{\phi}$, and $\hat{p}$

Asymptotically, MLEs such as $\hat{\phi}$, $\hat{S}$, and $\hat{p}$ are normally distributed. However, if the coefficients of variation of these parameter estimators are too large, the normal approximation is poor, especially for $\hat{S}$ and $\hat{\phi}$. Hypothesis testing and confidence intervals can be improved by using transformations that better approximate normality. Recommended transformations are the log-transform for $\hat{\phi}$ and $\hat{S}$ and the logistic (log-odds) transform for $\hat{p}$. These transformations are not routinely necessary; they make little difference if the cv of the parameter estimator in question is sufficiently small, say cv ≤ 0.1.

3.5.1. Log-Transform for $\hat{\phi}$ and $\hat{S}$

The MLEs of ϕ and S invariably have variance formulae expressible as

$$\text{var}(\hat{\phi}) = (\phi)^2 [\text{cv}(\hat{\phi})]^2,$$

or

$$\text{var}(\hat{S}) = (S)^2 [\text{cv}(\hat{S})]^2,$$

where the corresponding coefficients of variation are relatively stable as $\hat{\phi}$ or $\hat{S}$ varies. In contrast, there is a high correlation between $\hat{\phi}$ and $\hat{\text{var}}(\hat{\phi})$. The natural log-transformation greatly reduces this correlation. Asymptotically,

$$\hat{\text{var}}[\ln(\hat{\phi})] = [\text{cv}(\hat{\phi})]^2;$$

$$\hat{\text{var}}[\ln(\hat{S})] = [\text{cv}(\hat{S})]^2.$$

Experience based in part on simulation results has shown that the distributions of $\ln(\hat{\phi})$ and $\ln(\hat{S})$ are more nearly normal, especially when the cv of $\hat{\phi}$ or $\hat{S}$ is large. As a useful rule of thumb, we suggest that cv is small at $\leq 10\%$, moderate near 20%, and large at $\geq 40\%$. The normal approximation for the distributions of $\hat{\phi}$ or $\hat{S}$ is poor when a cv of $\hat{\phi}$ or $\hat{S}$ is large. In fact, if

a cv is large, the asymptotic variance formula itself is not good. However, if some random variable x is really log-normal, the exact variance of $\ln(x)$ is

$$\text{var}[\ln(x)] = \ln(1 + [\text{cv}(x)]^2) .$$

Consequently, if the cv of some survival or treatment effect parameter estimator is not small, we recommend computing tests and confidence intervals based on the log-transform with

$$\hat{\text{var}}[\ln(\hat{\phi})] = \ln(1 + [\text{cv}(\hat{\phi})]^2) , \text{ and}$$

$$\hat{\text{var}}[\ln(\hat{S})] = \ln(1 + [\text{cv}(\hat{S})]^2) .$$

Manly (1984) gave a confidence interval procedure for ϕ that is specific to the Jolly-Seber model. The above method based on $\ln(\phi)$ is much simpler than Manly's method, yet performs almost as well (unpublished investigations of the authors).

Hypothesis tests of the type H_0: $\phi_{t1} = \phi_{c1}$ based on $\hat{\phi}_{t1} - \hat{\phi}_{c1}$ are less sensitive to the need for a transform. However, if one wants to test the equivalent hypothesis H_0: $S = 1$ based on $\hat{S}$, we recommend the log-transformation unless the $\text{cv}(\hat{S})$ is small. The corresponding log-based confidence intervals are also recommended. For example, for an approximate $(1 - \alpha)100\%$ CI, one computes lower and upper bounds, $\hat{S}_L$ and $\hat{S}_U$, as

$$\hat{S}_L = \hat{S}/C$$

and

$$\hat{S}_U = \hat{S}C ,$$

where

$$C = \exp\left[z_{\alpha/2}\sqrt{\ln(1 + [\text{cv}(\hat{S})]^2)}\right] .$$

This approach still does not solve the problem of $\hat{S}_U$ being possibly greater than one, but that problem cannot be solved until one constrains $\hat{S} \leq 1$. Moreover, in many experiments there will be no logical reason to constrain S either ≤ 1 or ≥ 1; thus, we have not pursued such a constraint here.

Asymptotic covariances are also simple for log-transformed $\hat{\phi}$ and $\hat{S}$:

$$\text{cov}[\ln(\hat{\phi}_t), \ln(\hat{\phi}_c)] = \frac{\text{cov}(\hat{\phi}_t, \hat{\phi}_c)}{\phi_t \phi_c}.$$

For inference purposes, we are assuming that $\hat{\phi}_t$ and $\hat{\phi}_c$ have a bivariate log-normal distribution. Then the exact relationship between these two covariances is

$$\text{cov}[\ln(\hat{\phi}_t), \ln(\hat{\phi}_c)] = \ln\left[1 + \frac{\text{cov}(\hat{\phi}_t, \hat{\phi}_c)}{\phi_t \phi_c}\right].$$

(This result is from Johnson and Kotz 1972:20.)

When one uses the log-transform, any $\hat{S}$ estimator of treatment effect becomes a linear function of $\hat{\phi}$-estimators. It is then easy to write the variance of $\ln(\hat{S})$ and the covariance between $\ln(\hat{S}_a)$ and $\ln(\hat{S}_b)$ for any pair of estimators $\hat{S}_a$ and $\hat{S}_b$. For example, if $\hat{S}_a = \hat{\phi}_{v1,i}/\hat{\phi}_{v2,i}$ and $\hat{S}_b = \hat{\phi}_{v3,j}/\hat{\phi}_{v4,j}$, then

$$\text{var}[\ln(\hat{S}_a)] = \ln\left(1 + [\text{cv}(\hat{\phi}_{v1,i})]^2\right) + \ln\left(1 + [\text{cv}(\hat{\phi}_{v2,i})]^2\right)$$

$$- 2\ln\left(1 + \frac{\text{cov}(\hat{\phi}_{v1,i}, \hat{\phi}_{v2,i})}{\phi_{v1,i}\, \phi_{v2,i}}\right).$$

Also,

$$\text{cov}[\ln(\hat{S}_a), \ln(\hat{S}_b)] = \ln\left(1 + \frac{\text{cov}(\hat{\phi}_{v1,i}, \hat{\phi}_{v3,j})}{\phi_{v1,i}\, \phi_{v3,j}}\right) - \ln\left(1 + \frac{\text{cov}(\hat{\phi}_{v1,i}, \hat{\phi}_{v4,j})}{\phi_{v1,i}\, \phi_{v4,j}}\right)$$

$$- \ln\left(1 + \frac{\text{cov}(\hat{\phi}_{v2,i}, \hat{\phi}_{v3,j})}{\phi_{v2,i}\phi_{v3,j}}\right) + \ln\left(1 + \frac{\text{cov}(\hat{\phi}_{v2,i}\hat{\phi}_{v4,j})}{\phi_{v2,i}\phi_{v4,j}}\right).$$

A case that might arise is $\hat{S} = (\hat{\phi}_{t1}\hat{\phi}_{t2})/(\hat{\phi}_{c1}\hat{\phi}_{c2})$; using the above results and standard linear statistical theory it is easy to write var[ln(S)].

In general, we recommend the variance and covariance formulae based on treating the $\hat{\phi}$ as log-normal random variables. When coefficients of variation are small, the confidence intervals based on a log-transformation are almost identical to the asymptotic results. However, as coefficients of variation increase, the log-normal distribution for $\hat{\phi}$ provided a better approximation than the assumption of a normal distribution.

3.5.2. Log-Odds Transform for $\hat{p}$

The desirability of a transform on $\hat{p}$ is most noticeable when one computes a confidence interval $\hat{p}_L$ to $\hat{p}_U$ and finds $\hat{p}_L < 0$ or $\hat{p}_U > 1$. Because $\hat{p}$ will always be in [0,1], we want a transformation that maps [0,1] to $(-\infty, +\infty)$. Two commonly used transformations in this case are $\arcsin\sqrt{\hat{p}}$ and $\ln[\hat{p}/(1-\hat{p})]$, of which we recommend the latter (log-odds or logistic transform). The corresponding variance formula is

$$\operatorname{var}\left[\ln[\hat{p}/(1-\hat{p})]\right] = \frac{\operatorname{se}(\hat{p})}{p(1-p)}.$$

Treating the transformed variable as normally distributed leads to an approximate $(1 - \alpha)$ 100% CI, as

$$\hat{p}_L = \frac{\hat{p}}{\hat{p} + (1-\hat{p})\,C}$$

and

$$\hat{p}_U = \frac{\hat{p}}{\hat{p} + (1-\hat{p})/C},$$

where

$$C = \exp\left[\frac{z_{\alpha/2}\operatorname{se}(\hat{p})}{\hat{p}(1-\hat{p})}\right].$$

3.6. Computing Theoretical Biases, Standard Errors, and Test Powers

There is a "quick and easy" numerical way to get a good idea of the bias of any estimator when the data do not fit the model. Assume that you postulate the parameters under model H_{2p} and want to know the biases that would occur in $\hat{\phi}_{t1}$ and $\hat{\phi}_{c1}$ (for example) if model $H_{1\phi}$ was used to analyze data arising from this specific case of model H_{2p}. Using the postulated values of ϕ_{v1} and p_{v2}, $v = 1, ..., V$ and $\phi_2, ..., \phi_{k-1}, p_3, ..., p_k$ for given k, and with specified R_{v1}, $v = 1, ..., V$, generate the expected values of m_{vij} and $R_{v2}, ..., R_{v,k-1}$, $v = 1, ..., V$. Next, analyze these expected data under $H_{1\phi}$ as if they were actual data (e.g., using an option called EXPECT in PROC SIMULATE of program RELEASE). The computed values of the $\hat{\phi}$, $\hat{p}$,

and $\hat{S}$ estimators are good approximations to the expected values of the same estimators when the given model H_{2p} holds but $H_{1\phi}$ is used for the analysis.

In the procedure presented above, expected or known losses on capture could be allowed. Another refinement is to specify the releases R_{vi}, $i = 1, ..., k - 1$, $v = 1, ..., V$ at each release site, not just the first, and then generate $E(m_{vij} \mid R_{vi})$ rather than $E(m_{vij} \mid R_{v1})$.

The general procedure is strictly numerical but is "analytic," not Monte Carlo. One just completely specifies any given model and protocol and then generates expected data under that model and protocol. The model used to generate the expected data is the true model. The expected data are then analyzed, as if they were real data, under any alternative model for the same protocol. This analysis produces approximate expected values of the ML estimators under the analysis model used, given the true model used to generate the (expected) data. This procedure is suitable for determining if the bias associated with using the alternative model, rather than the true model, is small, medium, or large (e.g., $\leq$2.5%, about 10%, or $\geq$20%).

Further information accrues from this numerical procedure. The standard errors (or variances) produced are the theoretical standard errors. One could use them to compute a standardized bias:

$$\Delta = \frac{\text{bias}(\hat{\theta})}{\text{se}(\hat{\theta})}$$

(for any parameter θ). If $\hat{\theta}$ is assumed to be normally distributed, Cochran (1963:14) showed the effect of Δ on confidence interval coverage. Basically, if $\Delta \leq 0.5$, the effect is negligible (see Chapter 5.2). Of particular interest would be to analyze the expected data under the true model generating those "data," as this produces theoretical standard errors under the assumed model.

Information on the power of individual tests can also be extracted. When an analysis (model) is applied to expected values, $E(m_{vij} \mid R_{v1})$, any chi-square test statistic produced is really the noncentrality parameter of the corresponding noncentral chi-square power curve of that test. Table 3.4 gives some power values versus the noncentrality parameter for several degrees of freedom. (With current software like SAS, it is easy to compute any noncentral chi-square distribution.) For example, we could specify an instance of model $H_{1\phi}$, in terms of releases R_{t1} and R_{c1}, $k = 6$, and parameters ϕ_{t1}, ϕ_{c1}, ϕ_2, ..., ϕ_{k-1}, p_2, ..., p_k. One could then produce $E(m_{tij} | R_{t1})$ and $E(m_{cij} | R_{c1})$, $i = 1, ..., k - 1, j = i + 1, ..., k$ and analyze these as real data. TEST 1 (model H_0 versus $H_{4\phi}$) has 9 df; TEST 1.*R*1 (model H_0 versus $H_{1\phi}$) has 1 df. If the computed values of TEST 1 and its component 1.*R*1 are 15.0 (theoretically the noncentrality parameters of TESTs 1.*T*2 through 1.*R*5 are zero when model $H_{1\phi}$ is true), then, from Table 3.4, the power of TEST 1 is about 0.75 while the power of TEST 1.*R*1 is about 0.97.

If rounding (to integers) is done before "data" analysis, the biases, standard errors, and noncentrality parameters will be affected slightly. However, results will still be useful for judging when model bias is a problem. Similarly, one will be able to tell if power is poor, medium,

or good. Attainment of the same level of precision regarding this information by Monte Carlo methods would probably require at least 1,000 replications of simulated data.

The validity of this theoretical evaluation of model bias, precision, and power depends on sample size being large. This procedure can give poor results if the R_{ui} are small. Monte Carlo methods are necessary to investigate small-sample properties of statistical procedures. Also, this numerical procedure does not aid in determining the properties of complex procedures such as model selection, which involves a sequence of steps. Finally, one cannot learn anything about the sampling distribution of estimators or statistics from this analytical procedure. Hence, there is still a need for Monte Carlo procedures (e.g., see Buckland 1984); however, simulation is not needed to determine asymptotic model bias, precision, or power.

Table 3.4. – Some powers for an α = 0.05-level chi-square test (i.e., under the null hypothesis that the test statistic has a central chi-square distribution) for selected df and a range of noncentrality parameter values.

	df				df			
Noncentrality parameter	1	2	5	Noncentrality parameter	10	20	30	40
0	0.05	0.05	0.05	0	0.05	0.05	0.05	0.05
1	0.17	0.13	0.10	2	0.12	0.10	0.09	0.08
2	0.29	0.23	0.16	4	0.21	0.16	0.13	0.12
3	0.41	0.32	0.22	6	0.32	0.23	0.19	0.17
4	0.52	0.42	0.29	8	0.43	0.31	0.26	0.22
5	0.61	0.50	0.36	10	0.54	0.40	0.33	0.28
6	0.69	0.58	0.43	12	0.64	0.49	0.40	0.35
7	0.75	0.66	0.50	14	0.72	0.57	0.48	0.41
8	0.81	0.72	0.56	16	0.79	0.65	0.55	0.48
9	0.85	0.77	0.62	18	0.85	0.72	0.62	0.55
10	0.89	0.82	0.68	20	0.89	0.78	0.68	0.61
11	0.91	0.85	0.73	22	0.92	0.82	0.74	0.67
12	0.93	0.88	0.77	24	0.95	0.87	0.79	0.72
13	0.95	0.91	0.81	26	0.96	0.90	0.83	0.77
14	0.96	0.93	0.84	28	0.98	0.92	0.87	0.81
15	0.97	0.94	0.87	30	0.98	0.94	0.90	0.84
16	0.98	0.96	0.89	32	0.99	0.96	0.92	0.88
17	0.98	0.97	0.91	34	0.99	0.97	0.94	0.90
18	0.99	0.97	0.93	36	1.00	0.98	0.95	0.92
19	0.99	0.98	0.94	38	1.00	0.99	0.97	0.94
20	0.99	0.99	0.95	40	1.00	0.99	0.97	0.95

The justification for this analytical methodology comes from large-sample and maximum likelihood theory as regards bias and standard error evaluation. This method of test power evaluation is justified by more recent work, for example, Moore (1984). This theory provides further ways to achieve computing efficiency. Asymptotic model bias is independent of the numbers of fish released, R_{v1} (note that this theory also applies to Jolly-Seber models with releases at each occasion $i = 1, ..., k - 1$). Thus, with model bias (if releases are not few), one only needs to do the computations for one set of release numbers.

Let the parameter θ represent ϕ or p. The variance of $\hat{\theta}$ is proportional to the reciprocal of the release numbers, R_{v1}. Let the theoretical standard errors be evaluated numerically (as above) based on releases R_{v1}, $v = 1, ..., V$. In this situation, one can denote the standard error of $\hat{\theta}$ as $\text{se}(\hat{\theta} \mid R_{v1})$. Then, if all releases are multipled to be a constant, C, (e.g., C = 0.1 or 100), the standard error under these alternate releases, all else being the same, is

$$\text{se}(\hat{\theta} \mid CR_{v1}) = \frac{\text{se}(\hat{\theta} \mid R_{v1})}{\sqrt{C}}.$$

Therefore, one could set all $R_{v1} = 100{,}000$, compute analytical results, and determine standard errors if releases were, for example, 1,000 by using the above relationship with C = 0.01 = 1,000/100,000. (The advantage of using $R_{v1} = 100{,}000$ is the minimization of rounding error when expected captures are rounded to integers.) Once $\text{se}(\hat{\theta})$ is known for any releases R_{v1}, it is essentially known for all releases.

Similar computing economics occur for evaluating test power. Let δ represent the noncentrality parameter of any test. The value of δ depends on the releases and other factors (i.e., the true model, the value of k, the ϕ_{vi}, and the p_{vi}). Denote the dependence of δ on release numbers by $\delta(R_{v1})$. Then if the release numbers are CR_{v1}, everything else being the same, one has

$$\delta(CR_{v1}) = C\delta(R_{v1}).$$

For example, doubling the number of released fish doubles the noncentrality parameter. As a consequence of these analytical properties of $\text{se}(\hat{\theta} \mid R_{v1})$ and $\delta(R_{v1})$, it is easy to evaluate the effect of simply altering the numbers of releases with regards to standard errors and test power.

3.7. Testing Losses on Capture for a Treatment Effect

For each recapture datum m_{vij}, there corresponds d_{vij} losses on capture. For example, one might have $m_{t13} = 500$, $d_{t13} = 17$ and $m_{c13} = 561$, $d_{c13} = 21$; then, a 2×2 contingency table is used to test whether the loss rate is the same for treatments as for controls. In general, a $V \times 2$ contingency table is used for this testing for each $i = 1, ..., k - 1, j = i + 1, ..., k$. Losses on capture often are few or nil; such testing is then not needed.

Under our multinomial modeling approach, we obtain the conditional distributions

$$d_{vij} \mid m_{vij} \sim \text{bin}(m_{vij}; \gamma_{vij})$$

independently for $v = 1, ..., V$ and all i, j. Under the hypothesis H_0: $\gamma_{vij} = \gamma_{ij}$, one has the distribution

$$\Pr\{d_{1ij}, ..., d_{vij} \mid H_0\} = \frac{\prod_{v=1}^{V} \binom{m_{vij}}{d_{vij}}}{\binom{m_{.ij}}{d_{.ij}}}.$$

If data are not too sparse, one can use the $V \times 2$ contingency table chi-square test of homogeneity. For sparse data, a useful ad hoc procedure is to pool some of the tables, say, over j. Computing and examining the ratios d_{vij}/m_{vij} could also be useful.

In the case of $V = 2$, one can conveniently examine the one-sided alternative H_A: a higher (or lower) loss on capture rate for treatment fish. An example is the use of the test statistic

$$z = \frac{\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \left[\left(d_{tij}/m_{tij}\right) - \left(d_{cij}/m_{cij}\right)\right]}{\sum_{i=1}^{k-1} \sum_{j=i+1}^{k} \sum_{v=c}^{t} \frac{1}{m_{vij}} \left(\frac{d_{vij}}{m_{vij}}\right) \left(1 - \frac{d_{vij}}{m_{vij}}\right)},$$

where z is approximately a standard normal variable under the null hypothesis. The point here is that one can test for a treatment effect of losses on capture. This testing reduces to a standard statistical problem of examining proportions; numerous statistical methods are available for this situation (see Fleiss 1981).

3.8. Handling Effects

One reason we use the Jolly-Seber model as the starting point for developing a general theory for each group is the nonidentifiability problems that arise with other more general models. The first generalization of the Jolly-Seber model that arises is to allow for a "release" or handling effect for fish recaptured then rereleased at dams 2, ..., k - 1. In particular, survival, ϕ_{vj}, after capture, handling, and release at dam j is likely to be affected. Such a handling effect is well known in the literature (see, for example, Manly 1971a; Brownie and Robson 1983; Arnason and Mills 1986). In general, just the capturing and handling may affect

subsequent survival after release. In addition, there may be a capture and release effect in stud-ies at hydroelectric dams. If the fish are released below the dam (or even above the dam), they will probably experience different mortality stresses at that dam compared to fish passing the dam but not caught there. (This problem has implications for system-wide studies.)

We can generalize the Jolly-Seber model by allowing such released fish at dam j to have survival rate ϕ'_j between dams j and $j+1$; after that, their survival rate is the same as that of other fish in the study (a one-period effect). It suffices to deal only with one treatment group here so we drop the subscript v. Now the parameters of the model are $\phi'_1, ..., \phi'_{k-1}, \phi_2, ..., \phi_{k-1}$ and $p_2, ..., p_k$. Note that we use ϕ'_1 here, not ϕ_1, for consistency of notation; we cannot define a handling effect at first release. When $\phi'_j \neq \phi_j$, $j = 2, ..., k - 1$, none of these parameters are identifiable. Moreover, none of our tests of hypotheses can detect such a handling effect. Thus, there is no point in considering such a model, even though it may be real. Considering more general models may be equally pointless unless either they entail getting additional information or parameters are identifiable. We here give the mathematical basis supporting this lack of identifiability.

Given the R_i, the $m_{i,i+1}, ..., m_{ik}, R_i - r_i$ are independent multinomials. The model is then (essentially) the structure we put on

$$\frac{E(m_{ij} \mid R_i)}{R_i} = \pi_{ij}.$$

For Jolly-Seber, a standardized representation is

$$\pi_{ij} = \begin{cases} \alpha_i, & j = i + 1 \\ \beta_i \cdots \beta_{j-2}\alpha_{j-1}, & j > i + 1, \end{cases}$$

for $i = 1, ..., k - 1$, $j = i + 1, ..., k$ with $\alpha_i = \phi_i p_{i+1}$ and $\beta_i = \phi_i q_{i+1}$. All that matters is the structure of these π_{ij}, i.e., for Jolly-Seber, they are time-specific only.

Under the one-period capture-handling and release-effects model superimposed on Jolly-Seber, we have the model structure as, for example, when $k = 5$

i	$k = 2$	3	4	5
1	$(\phi'_1 p_2)$	$(\phi'_1 q_2)(\phi_2 p_3)$	$(\phi'_1 q_2)(\phi_2 q_3)(\phi_3 p_4)$	$(\phi'_1 q_2)(\phi_2 q_3)(\phi_3 q_4)(\phi_4 p_5)$
2		$(\phi'_2 p_3)$	$(\phi'_2 q_3)(\phi_3 p_4)$	$(\phi'_2 q_3)(\phi_3 q_4)(\phi_4 p_5)$
3			$(\phi'_3 p_4)$	$(\phi'_3 q_4)(\phi_4 p_5)$
4				$(\phi'_4 p_5)$

The structure of these π_{ij} can be rearranged to arrive at the following standardized representation:

i	$k = 2$	3	4	5
1	α'_1	$\beta'_1\alpha'_2$	$\beta'_1\beta'_2\alpha'_3$	$\beta'_1\beta'_2\beta'_3\alpha'_4$
2		α'_2	$\beta'_2\alpha'_3$	$\beta'_2\beta'_3\alpha'_4$
3			α'_3	$\beta'_3\alpha'_4$
4				α'_4

Here,

$$\alpha'_i = \phi'_i p_{i+1}, \quad i = 1, ..., k - 1;$$

$$\beta'_i = (\phi'_i q_{i+1}\phi_{i+1} / \phi'_{i+1}), \quad i = 1, ..., k - 2.$$

This form is the same as Jolly-Seber. It therefore has the same MSS and cannot be distinguished from Jolly-Seber; i.e., we cannot test H_0: $\phi_i = \phi'_i$. Individual parameters are not estimable under this model (i.e., the ϕ'_i, ϕ_i, p_i). Intrinsically estimable are the α'_i and β'_i ($2k$ - 3 parameters).

Only straightforward algebra is needed to derive these results. For example,

$$\pi_{13} = \phi'_1 q_2 \phi_2 p_3$$

$$= \phi'_1 q_2 \frac{\phi_2}{\phi'_2} \phi'_2 p_3$$

$$= \beta'_1 \alpha'_2 .$$

If one considers the large-sample expected values of the usual Jolly-Seber estimators of ϕ_i and p_i, one realizes that further algebraic manipulation is possible. We end up with the representation

$$\alpha'_i = \phi^*_i p^*_{i+1}, \quad i = 1, ..., k - 1;$$

$$\beta'_i = \phi^*_i q^*_{i+1}, \quad i = 1, ..., k - 2;$$

where

$$\phi^*_i = \phi'_i\left[p_{i+1} + q_{i+1}\frac{\phi_{i+1}}{\phi'_{i+1}}\right], \quad i = 1, ..., k-1;$$

$$p^*_i = \frac{p_i}{p_i + q_i\dfrac{\phi_i}{\phi'_i}}, \quad i = 2, ..., k.$$

This shows that if we generalize the Jolly-Seber model to allow a first-period effect on survival rate after release on occasion j, then the model structure of the applicable π_{ij} is identical to

$$\pi_{ij} = \begin{cases} (\phi^*_i p^*_{i+1}), & j = i+1 \\ (\phi^*_i q^*_{i+1}) \cdots (\phi^*_{i-2} q^*_{i-1})(\phi^*_{i-1} p^*_i), & j > i+1 \end{cases},$$

with ϕ^*_i and p^*_i as given above. This structure is exactly time-specific, and thus is a Jolly-Seber model. We will not be able to detect a release and handling effect by any tests applied to the usual release-recapture data. However, the Jolly-Seber estimators are badly affected; their expected values are the above ϕ^*_i and p^*_i.

We note that, under $H_{1\phi}$, the estimator of $S = \phi_{t1}/\phi_{c1}$ is unaffected by this problem ($\phi_{c1} - \phi_{t1}$ *is* affected). However, under H_{2p}, $\hat{S}$ is affected; the expected value of $\hat{S}$ is then

$$E(\hat{S}) = S\,\frac{\left[p_{t2} + q_{t2}\dfrac{\phi_2}{\phi'_2}\right]}{\left[p_{c2} + q_{c2}\dfrac{\phi_2}{\phi'_2}\right]}.$$

When the capture probabilities are small, the bias in $\hat{S}$ is small. Under model $H_{2\phi}$, however, a handling effect such as the one considered here can seriously bias even $\hat{S}$. Further work on this problem, such as its affect on $\hat{N}$ in the Jolly-Seber model, was given by Arnason and Mills (1986).

3.9. Bias Reduction by Peeling for First Capture History and Unknown Capture History Protocols

In Chapters 2.2 and 2.3, we mentioned that, if model $H_{1\phi}$ does not fit the data, then $\hat{S}$, computed as if $H_{1\phi}$ were true, will be biased. Under these protocols, no separate ϕ or p

parameters are estimable for $H_{1\phi}$ (or for any more general models). For the other protocols, the strategy to reduce model bias is to seek a more general model by using some specific tests. A version of that strategy is applicable for these two protocols; however, it is less rigorous and less satisfactory. In the case of the first capture history protocol in particular, bias is not eliminated, it is only reduced. We next give equations for approximate bias evaluation relative to this peeling strategy.

Two concepts of bias must be distinguished: (1) statistical bias and (2) model bias. If $\hat{\theta}$ is some estimator and the assumptions (i.e., model) underlying $\hat{\theta}$ are true, we may still have $E(\hat{\theta}) \neq \theta$; thus, $\hat{\theta}$ is statistically biased. For example, even under $H_{1\phi}$,

$$\hat{S} = \frac{r_{t1}/R_{t1}}{r_{c1}/R_{c1}}$$

suffers some statistical bias (see Section 3.4.3). Much more severe bias is likely to occur if one's model is wrong. We take model bias to mean the value of $E(\hat{\theta}) - \theta$ when the assumptions (i.e., models) are wrong. One can often adjust for statistical bias. We will consider model bias of various statistically bias-adjusted estimators under first capture history and unknown capture history protocols in Sections 3.9.1 and 3.9.2.

3.9.1. Model Bias and Peeling for the First Capture History Protocol

Under $H_{1\phi}$, we have

$$S = S_1 = \frac{r_{t1} / R_{t1}}{(r_{c1} + 1) / (R_{c1} + 1)}$$

as virtually unbiased. (Note: We will use a subscript on S in this section to denote degree of peeling.) As discussed in Section 2.2.8., one may do some tests and conclude that model $H_{1\phi}$ does not fit; thus, S suffers model bias. If $H_{2\phi}$ is not rejected, one recourse in reducing this model bias is to use

$$S_2 = \frac{(r_{t1} - m_{t12})/R_{t1}}{(r_{c1} - m_{c12} + 1) / (R_{c2} + 1)}.$$

In general, we could consider using

$$S_j = \frac{\left(\sum_{n=j+1}^{k} m_{t1n}\right)/R_{t1}}{\left(\sum_{n=j+1}^{k} m_{c1n} + 1\right)/(R_{c1} + 1)}, \quad i = 1, \ldots, k - 1.$$

The nearly exact expected value of S_j is

$$E(S_j) = \frac{\left(\sum_{n=j+1}^{k} \pi_{tin}\right)/R_{t1}}{\left(\sum_{n=j+1}^{k} \pi_{cin}\right)/R_{c1}}.$$

The expected value is irrespective of the true model assumptions about the ϕ_{vi} and p_{vi}. For example,

$$E(S) = E(S_1) = \frac{\lambda_{t1}}{\lambda_{t2}} = \frac{\phi_{t1}(p_{t2} + q_{t2}\lambda_{t2})}{\phi_{c1}(p_{c2} + q_{c2}\lambda_{c2})},$$

and

$$E(S_2) = \frac{\phi_{t1}q_{t2}\lambda_{t2}}{\phi_{c1}q_{c2}\lambda_{c2}} = \left(\frac{\phi_{t1}\phi_{t2}}{\phi_{c1}\phi_{c2}}\right)\left(\frac{q_{t2}}{q_{c2}}\right)\left(\frac{p_{t3} + q_{t3}\lambda_{t3}}{p_{c3} + q_{c3}\lambda_{c3}}\right).$$

In general, if model $H_{j\phi}$ holds,

$$E(S_j) = S\left[\prod_{i=2}^{j+1} \frac{q_{ti}}{q_{ci}}\right], \quad j = 2, \ldots, k-1.$$

Note that under model $H'_{j\phi}$, S_j is (essentially) unbiased because $H'_{j\phi}$ *assumes* that $p_{ti} = p_{ci}$ for all $i = 2, \ldots, k$.

Under model $H_{1\phi}$, $\tilde{S}$ is unbiased. However, under any higher model, particularly H_{2p} or even $H_{2\phi}$, $\tilde{S}$ suffers model bias. TEST 1.*T*2 tests $\tau_{t2} = \tau_{c2}$ ($\tau_{vi} = p_{vi}/(p_{vi} + q_{vi}\lambda_{vi})$). Because the null hypothesis is false if either $p_{t2} \neq p_{c2}$ or $\phi_{t2} \neq \phi_{c2}$, at best the alternative hypothesis is model $H_{2\phi}$ (not H_{2p}). However, we can evaluate theoretically $E(S_1)$ and $E(S_2)$ under model H_{2p}:

$$E(S_1) = S\left(\frac{p_{t2} + q_{t2}\lambda_2}{p_{c2} + q_{c2}\lambda_2}\right);$$

$$E(S_2) = S\frac{q_{t2}}{q_{c2}}.$$

If the capture probabilities are small, then approximately $E(S_1) \doteq S(p_{t2} / p_{c2})$ while $E(S_2) \doteq S$. More precise results can be obtained by computing numerical examples. For example, if H_{2p} held with $p_{t2} = 0.03$, $p_{c2} = 0.05$, and $\lambda_2 = 0.12$, then $E(S_1) = S(0.893)$, an 11% relative bias, whereas $E(S_2) = S(1.021)$, a 2% relative bias.

The simplest and perhaps most important alternative case is that H_{2p} might be true, not $H_{1\phi}$. One tests for this case (with first capture history data) by evaluating TEST 1.*T*2 and the sum of TESTS 1.*T*3 to 1.*Tk* - 1. If TEST 1.*T*2 rejects and the remaining tests do not, it is highly likely that S_1 will have substantial bias relative to its standard error. If the capture probabilities are known to be small, then we recommend peeling (discarding) m_{v12} because S_2 will be substantially less biased than S_1. (Note that matters are different with moderate or large capture probabilities; then, peeling with the above formulae could actually make bias worse.)

The following equation may provide an improved peeled estimator.

$$\tilde{\tilde{S}}_j = \frac{\left(\sum_{n=j+1}^{k} m_{t1n}\right) / \left(R_{t1} - \sum_{n=2}^{j} m_{t1n}\right)}{\left[\left(\sum_{n=j+1}^{k} m_{c1n}\right) + 1\right] / \left(R_{c1} + 1 - \sum_{n=2}^{j} m_{c1n}\right)}, \quad j = 2, \ldots, k - 1.$$

The almost exact expectation of $\tilde{\tilde{S}}_j$ is

$$E(\tilde{\tilde{S}}_j) = \frac{\left(\sum_{n=j+1}^{k} \pi_{t1n}\right) / \left(1 - \sum_{n=2}^{j} \pi_{t1n}\right)}{\left(\sum_{n=j+1}^{k} \pi_{c1n}\right) / \left(1 - \sum_{n=2}^{j} \pi_{c1n}\right)}, \quad j = 2, \ldots, k - 1.$$

In particular, under $H_{2\phi}$,

$$E(S_2) = S\left(\frac{1 - p_{t2}}{1 - \phi_{t1} p_{t2}}\right) / \left(\frac{1 - p_{c2}}{1 - \phi_{c1} p_{c2}}\right);$$

here $S = \phi_{t1}\phi_{t2}/(\phi_{c1}\phi_{c2})$ but this is the relevant treatment effect parameter. (If H_{2p} is true rather than $H_{2\phi}$, then $\phi_{t2} = \phi_{c2}$.) For small capture probabilities, the above formula for $E(S_2)$ is well approximated by

$$E(\tilde{\tilde{S}}_2) \doteq S\left(\frac{1 - p_{t2} + \phi_{t1}p_{t2}}{1 - p_{c2} + \phi_{c1}p_{c2}}\right),$$

which is closer to S than is $E(\tilde{S}_2)$.

3.9.2. Model Bias and Peeling for the Unknown Capture History Protocol

We draw on notation and theory presented in Chapter 2.3 in general and Section 2.3.5 in particular to obtain results. Under $H_{1\phi}$ the appropriate estimator of S (adjusted for statistical bias) is

$$\tilde{S} = \tilde{S}_1 = \frac{m_{t.}/R_{t1}}{(m_{c.} + 1)/(R_{c1} + 1)}.$$

Peeled estimators are

$$\tilde{S}_j = \frac{\left(\sum_{n=j+1}^{k} m_{tn}\right)/R_{t1}}{\left(\left(\sum_{n=j+1}^{k} m_{cn}\right) + 1\right)/(R_{c1} + 1)}, \quad j = 2, ..., k - 1.$$

Here, m_{vj} is the number of fish captured at dam j and $m_{v.} = m_{v2} + \cdots + m_{vk}$.

By drawing on results presented in Section 2.3.5, one can develop a formula for $E(\tilde{S}_j)$ in the general case of losses on capture. We present results here only for the case of no losses on capture; then

$$E(m_{vj}) = R_{v1}\phi_{v1} \cdots \phi_{v,j-1}p_{vj}, \quad j = 2, ..., k.$$

We define $f_{vj} = \phi_{v1} \cdots \phi_{v,j-1}p_{vj}$, $j = 2, ..., k$ and thus write the approximate expectations of the $\tilde{S}_j$ as

$$E(\tilde{S}_j) = \frac{\sum_{n=j+1}^{k} f_{tn}}{\sum_{n=j+1}^{k} f_{cn}}, \quad j = 1, ..., k - 1.$$

Under model $H_{j\phi}$ for this protocol,

$$E(\tilde{S}_j) = \frac{\phi_{t1} \cdots \phi_{tj}}{\phi_{c1} \cdots \phi_{cj}} = S .$$

Heuristically, it is as if dam $j + 1$ were actually the second dam and $H_{1\phi}$ was true.

One can evaluate the (approximate) bias to be expected in $\tilde{S}_1$ if, for example, H_{2p} is the true model:

$$E(\tilde{S}_1) = S\left(\frac{p_{t2} + Q}{p_{c2} + Q}\right);$$

$$Q = \phi_2 p_3 + \phi_2\phi_3 p_4 + \cdots + \phi_2 \cdots \phi_{k-1} p_k .$$

For example, if $p_{t2} = 0.03$, $p_{c2} = 0.05$, $Q = 0.124$ (which would arise from $k = 6$, $\phi_2 = \phi_3 = \phi_4 = \phi_5 = 0.9$, and $p_3 = p_4 = p_5 = p_6 = 0.04$), and H_{2p} is true, then $E(\tilde{S}_1) = S(0.885)$.

Clearly, this approach can be used to assess model bias under any scenario one can specify; thus, it evaluates the effect of peeling on bias.

3.10. Synthetic Example of Multiple Treatments

We present here a simulated example of data having three treatment levels and a control; thus, there are $V = 4$ groups. The partial capture history scheme B protocol is used. Table 3.5 shows the values of the parameters used in generating the sample data. At release point (time) 1, three treatment groups and a control are released ($R_v = 1{,}000$ each group). These releases are followed by five recapture occasions. We envision batch marks being used, as per scheme B. These marks would distinguish fish by lot and treatment. At recapture time 2, another mark is applied to all recaptures and those fish are released. At times 3, 4, 5, and 6, all captured fish are removed. The full simulated example comprises 10 lots of the four groups. Here we present some key results for lot 1; in Part 4 we use results from all 10 lots to illustrate empirical replication in a complex design.

Model H_{2p} is the true underlying model; thus, the four ϕ_{v1} and the four p_{v2} differ by treatment. No other parameters depend on treatment, however. Of particular interest are the treatment effects. With $v = 4$, there are six possible combinations of $S_{v,v'} = \phi_{v1}/\phi_{v'1}$. We simplify matters by saying that the only meaningful effects evolve from comparisons of each treatment separately with the single control. Thus, treatment effects are S_1, S_2, and S_3 defined by ϕ_{v1}/ϕ_{c1}, $v = 1, 2, 3$, with notation $v = 1$ for t_1, $v = 2$ for t_2, $v = 3$ for t_3, and $v = 4$ for c.

Table 3.5. – Parameter values used in the simulation example of three treatment levels and a single control (four groups), partial capture history scheme B data with $k = 6$ release-recapture sites. This is model H_{2p}. Ten lots were generated. Releases at site 1 are $R_{v1} = 1{,}000$ for each group and lot.

Survival parameters, treatment effects	Capture parameters
$\phi_{t1,1} = 0.81$, $S_1 = 0.90$	$p_{t1,2} = 0.1$
$\phi_{t2,1} = 0.675$, $S_2 = 0.75$	$p_{t2,2} = 0.1$
$\phi_{t3,1} = 0.63$, $S_3 = 0.70$	$p_{t3,2} = 0.1$
$\phi_{c,1} = 0.90$	$p_{c,2} = 0.2$
Common to all groups	**Common to all groups**
$\phi_2 = 0.85$	$p_3 = 0.2$
$\phi_3 = 0.80$	$p_4 = 0.2$
$\phi_4 = 0.70$	$p_5 = 0.2$
$\phi_5 = 0.85$	$p_6 = 0.2$

Table 3.6 shows what the input data look like in capture history matrix form. Note that, for most capture histories, the counts are shown as negative. For example, for $h = \{100010\}$, $X_{1h} = -38$, $X_{2h} = -31$, $X_{3h} = -29$, and $X_{4h} = -43$. At capture site 5, none of these fish were returned to the study. Fish were released only at sites 1 and 2. For example, the 10 fish for X_{1h}, $h = \{110100\}$ were released at site 1, recaptured and released at site 2, and recaptured at site 4 but not rereleased.

The reduced m-array representation of the data is constructed by RELEASE and is also shown in Table 3.6. Under partial capture history scheme B, there are no multiple subcohorts. Consequently, no components of TEST 3 exist.

TEST 2.*C*2 exists for each group. For example, for treatment group $v = 1$, TEST 2.*C*2 is computed from the 2×4 table

123	76	38	23
11	10	2	5

The results of this goodness of fit testing are given in Table 3.7. TEST 2.*C*2 for group 1 produces the chi-square value of 3.83 (3 df) and observed significance level $P = 0.280$. Summed over all four groups, we have the TEST 2 chi-square result of 12.36 (11 df), $P = 0.730$. Clearly, there is no reason to reject the basic Jolly-Seber model here (i.e., within each separate group, parameters are time-specific only).

Table 3.6. – The input capture history matrix, as printed by RELEASE, and the reduced *m*-array summaries of the data, by treatment group, for lot 1.

```
INPUT --- PROC CHMATRIX OCCASIONS=6 GROUPS=4 LOTS=10    ;

INPUT ---    GLABEL(1)=Treatment Group 1;
INPUT ---    GLABEL(2)=Treatment Group 2;
INPUT ---    GLABEL(3)=Treatment Group 3;
INPUT ---    GLABEL(4)=Control Group;

INPUT ---       LOT=1;
INPUT ---       100000    656   685   719   549;
INPUT ---       100010    -38   -31   -29   -43;
INPUT ---       101000   -123  -108  -100  -115;
INPUT ---       100100    -76   -80   -68   -78;
INPUT ---       110000     56    43    47   128;
INPUT ---       100001    -23   -27   -18   -25;
INPUT ---       110100    -10    -7    -3   -15;
INPUT ---       111000    -11   -10    -9   -27;
INPUT ---       110010     -2    -7    -5   -13;
INPUT ---       110001     -5    -2    -2    -7;


 Number of capture histories read was   10


                  Observed Recaptures for Group 1
                        Treatment Group 1

   i   R(i)          m(i,j)                    r(i)
              j=  2    3    4    5    6
   1   1000     84  123   76   38   23    344
   2     84         11   10    2    5     28


 m(j)           84  134   86   40   28
 z(j)          260  154   68   28    0
```

Table 3.6. – Continued.

Observed Recaptures for Group 2
Treatment Group 2

i	R(i)	m(i,j)				r(i)
	j= 2	3	4	5	6	
1	1000 69	108	80	31	27	315
2	69	10	7	7	2	26
m(j)	69	118	87	38	29	
z(j)	246	154	67	29	0	

Observed Recaptures for Group 3
Treatment Group 3

i	R(i)	m(i,j)				r(i)
	j= 2	3	4	5	6	
1	1000 66	100	68	29	18	281
2	66	9	3	5	2	19
m(j)	66	109	71	34	20	
z(j)	215	125	54	20	0	

Observed Recaptures for Group 4
Control Group

i	R(i)	m(i,j)				r(i)
	j= 2	3	4	5	6	
1	1000 190	115	78	43	25	451
2	190	27	15	13	7	62
m(j)	190	142	93	56	32	
z(j)	261	181	88	32	0	

Sums for the above Groups

m.	0	409	503	337	168	109
z.	0	982	614	277	109	
R.	4000	409	0	0	0	
r.	1391	135	0	0	0	

Table 3.7. – Results of TEST 2, goodness of fit, for the simulated model H_{2p}, scheme B data for lot 1. Only TEST 2.C2 exists under scheme B.

Treatment Group 1

TEST 2.C2: Test of row 1 vs. row 2

O	123	76	38	23	260
E	121.0	77.6	36.1	25.3	
C	0.0	0.0	0.1	0.2	
O	11	10	2	5	28
E	13.0	8.4	3.9	2.7	
C	0.3	0.3	0.9	1.9	
	134	86	40	28	288

Chi-square=3.8329 (df=3) P=0.2801

Treatment Group 2

TEST 2.C2: Test of row 1 vs. row 2

O	108	80	31	27	246
E	106.7	78.7	34.4	26.2	
C	0.0	0.0	0.3	0.0	
O	10	7	7	2	26
E	11.3	8.3	3.6	2.8	
C	0.1	0.2	3.1	0.2	
	118	87	38	29	272

Chi-square=4.0808 (df=3) P=0.2529

Treatment Group 3

TEST 2.C2: Test of row 1 vs. row 2

O	100	68	47	215
E	100.1	65.2	49.6	
C	0.0	0.1	0.1	
O	9	3	7	19
E	8.9	5.8	4.4	
C	0.0	1.3	1.6	
	109	71	54	234

Chi-square=3.1440 (df=2) P=0.2076

Table 3.7. – Continued.

```
                    Control Group
          TEST 2.C2:  Test of row 1 vs. row 2
              +------+------+------+------+
            O| 115  |  78  |  43  |  25  | 261
            E| 114.7| 75.1 | 45.3 | 25.9 |
            C|  0.0 |  0.1 |  0.1 |  0.0 |
             +------+------+------+------+
            O|  27  |  15  |  13  |  7   |  62
            E|  27.3| 17.9 | 10.7 |  6.1 |
            C|  0.0 |  0.5 |  0.5 |  0.1 |
             +------+------+------+------+
               142     93     56     32    323
           Chi-square=1.2981 (df=3) P=0.7296
```

Summary of TEST 2 (Goodness of fit) Results					
Group	Component	Chi-square	df	P-level	Sufficient Data
1	2.C2	3.8329	3	0.2801	Yes
2	2.C2	4.0808	3	0.2529	Yes
3	2.C2	3.1440	2	0.2076	Yes
4	2.C2	1.2981	3	0.7296	Yes
-	TEST 2	12.3558	11	0.3375	-

Next we look at the results of TEST 1, summarized in Table 3.8. Not all possible components of TEST 1 can be computed for scheme B data; only 1.*R*1, 1.*R*2, 1.*T*2, 1.*T*3, 1.*T*4, and 1.*T*5 components exist. Each component is (here) based on a 2×4 table of summary statistics. The full contingency table for TEST 1.*T*2 is given in Table 3.8. The rows of that contingency table are m_{v2}, z_{v2} (which sum to T_{v2}) for $v = 1, 2, 3, 4$. The summary statistics m_{v2} and z_{v2} are given in Table 3.6. For example, $m_{v2} = 84, z_{v2} = 260$. The body of the contingency table for TEST 1.*T*2 is thus easily found to be

	v	m	z
	1	84	260
	2	69	246
	3	66	215
(control)	4	190	261

Table 3.8. – Some results of TEST 1, model selection, for the simulated model H_{2p}, scheme B data for lot 1. Only TESTs 1.*R*1, 1.*R*2, 1.*T*2, 1.*T*3, 1.*T*4, and 1.*T*5 exist here.

```
TEST 1.T2:  Test of p(2) equal across groups,
assuming higher order parameters are equal across groups.
              +------+------+
             O|  84  | 260  | 344
             E| 101.1| 242.9|
             C|   2.9|   1.2|
              +------+------+
             O|  69  | 246  | 315
             E|  92.6| 222.4|
             C|   6.0|   2.5|
              +------+------+
             O|  66  | 215  | 281
             E|  82.6| 198.4|
             C|   3.3|   1.4|
              +------+------+
             O| 190  | 261  | 451
             E| 132.6: 318.4|
             C|  24.8|  10.3|
              +------+------+
               409    982   1391
      Chi-square=52.5708 (df=3) P=0.0000

TEST 1.R1:  Test of Phi(1) equal across groups,
assuming higher order parameters are equal across groups.
              +------+------+
             O| 344  | 656  |1000
             E| 347.7| 652.2|
             C|   0.0|   0.0|
              +------+------+
             O| 315  | 685  |1000
             E| 347.7| 652.2|
             C|   3.1|   1.6|
              +------+------+
             O| 281  | 719  |1000
             E| 347.7| 652.2|
             C|  30.7|  16.3|
              +------+------+
              1391   2609   4000
      Chi-square=71.4344 (df=3) P=0.0000
```

Table 3.8. – Continued.

Summary of TEST 1 (Between Groups Test) Results

Component	Chi-square	df	P-level	Sufficient Data
1.T5	0.9793	3	0.8063	Yes
1.T4	1.3006	3	0.7290	Yes
1.T3	0.9354	3	0.8169	Yes
1.R2	1.2292	3	0.7460	Yes
1.T2	52.5708	3	0.0000	Yes
1.R1	71.4344	3	0.0000	Yes
TEST 1	128.4498	18	0.0000	

The chi-square test statistic for TEST 1.*T*2 is 52.5708 (3 df); we reject the null hypothesis that model $H_{1\phi}$ fits these data. The result for TEST 1.*R*1 reinforces this conclusion. Further examination of the summary results in Table 3.8 show that TEST 1 components 1.*R*2 and 1.*T*2 through 1.*T*5 do not reject. The sum of the corresponding chi-squares is 4.4445 (12 df), which also supports the conclusion that model H_{2p} provides an appropriate model for these data.

Table 3.9 shows for lot 1 the parameter estimates under model H_{2p}. Under model H_{2p}, scheme B, the only estimable parameters of interest are ϕ_{v1} and p_{v2}, $v = 1, ..., V$. For treatment group 1, from Table 3.9, we have

$$\hat{\phi}_{11} = 0.872, \hat{\text{se}}(\hat{\phi}_{11}) = 0.069;$$

$$\hat{p}_{12} = 0.096, \hat{\text{se}}(\hat{p}_{12}) = 0.013.$$

Also printed out are the asymptotic 95% confidence limits on the true parameter. For example, for ϕ_{11}, those limits are $0.872 \pm 1.96\,(0.069)$, or 0.736 to 1.007. The true value of ϕ_{11} is 0.81 and $p_{12} = 0.1$. For the control group ($v = 4$ here), we have

$$\hat{\phi}_{41} = 0.98, \hat{\text{se}}(\hat{\phi}_{41}) = 0.069.$$

Table 3.9. – Output from program RELEASE for model H_{2p} for the simulated data: scheme B, k = 6, 4 groups, lot 1.

Maximum Likelihood Estimates under Model H2p

Parameter	Estimate	Standard Error	95% Confidence Intervals Lower	Upper
		Estimates for Group 1		
		Treatment Group		
Phi(1)	0.871704	0.069207	0.736057	1.007350
p(2)	0.096363	0.012532	0.071800	0.120926
		Estimates for Group 2		
		Control Group		
Phi(1)	0.814289	0.066485	0.683978	0.944600
p(2)	0.084737	0.011894	0.061424	0.108049
		Estimates for Group 3		
Phi(1)	0.717370	0.060252	0.599277	0.835464
p(2)	0.092003	0.013146	0.066236	0.117770
		Estimates for Group 4		
Phi(1)	0.980733	0.068750	0.845983	1.115484
p(2)	0.193733	0.018519	0.157434	0.230031

Ratio of Survivals between Groups

Parameter	Estimate	Standard Error	95% Confidence Intervals Lower	Upper
S(1,2,Phi(1))	1.070509	0.073867	0.925730	1.215288
Corr(1,2,Phi(1))		0.633146		
S(1,3,Phi(1))	1.215138	0.087744	1.043159	1.387117
Corr(1,3,Phi(1))		0.610609		
S(1,4,Phi(1))	0.888828	0.056119	0.778835	0.998822
Corr(1,4,Phi(1))		0.649622		
S(2,3,Phi(1))	1.135102	0.083971	0.970519	1.299686
Corr(2,3,Phi(1))		0.601384		
S(2,4,Phi(1))	0.830286	0.054170	0.724113	0.936458
Corr(2,4,Phi(1))		0.639807		
S(3,4,Phi(1))	0.731463	0.050160	0.633150	0.829777
Corr(3,4,Phi(1))		0.617033		

It is worth noting the formula and computation of these $\hat{\phi}_{v1}$ and $\hat{p}_{v2}$:

$$\hat{\phi}_{v1} = \frac{r_{v1}}{R_{v1}}\left[\frac{m_{v2}}{T_{v2}} + \frac{z_{v2}R_{.2}}{T_{v2}r_{.2}}\right];$$

$$\hat{p}_{v2} = \frac{m_{v2}}{m_{v2} + \dfrac{z_{v2}R_{.2}}{r_{.2}}}.$$

From Table 3.6, one finds the sums $r_{.2} = 135$ and $R_{.2} = 409$, and the control group,

$$r_{41} = 451;$$

$$R_{41} = 1{,}000;$$

$$m_{42} = 190;$$

$$z_{42} = 261;$$

(hence $T_{42} = 451$). Thus, for example,

$$\hat{p}_{42} = \frac{190}{190 + \dfrac{261 \times 409}{135}}$$

$$= 0.194.$$

Finally, we get to the estimates of treatment effect on survival (also shown in Table 3.9). RELEASE automatically computes all such possible treatment effects, i.e.,

$$\hat{S}_{v,v'} = \frac{\hat{\phi}_{v1}}{\hat{\phi}_{v'1}}.$$

For example, S(1, 4, Phi(1)) denotes $\hat{S}_{1,4} = \hat{\phi}_{t1,1}/\hat{\phi}_{c1}$. One should examine only the effects that are meaningful. In this example,

$$\hat{S}_{14} = \hat{S}_1 = 0.889 \text{ (se} = 0.056);$$

$$\hat{S}_{24} = \hat{S}_2 = 0.830 \text{ (se} = 0.054);$$

$$\hat{S}_{34} = \hat{S}_3 = 0.731 \text{ (se} = 0.050).$$

Note the true values are $S_1 = 0.90$, $S_2 = 0.75$, and $S_3 = 0.70$. The approximate 95% CIs are also shown in Table 3.9 for each estimate of S_i. For example, for S_2, that interval is 0.724 to 0.936, which includes the true value even though $\hat{S}_2 = 0.830$.

Along with the ratios of $\hat{\phi}_{vi}$, RELEASE shows the estimated sampling correlations between the $\hat{\phi}_{vi}$ and $\hat{\phi}_{v'i}$ used in $\hat{S}_{v,v'} = \hat{\phi}_{vi}/\hat{\phi}_{v'i}$. These sampling correlations are labeled

Corr(v, v′, Phi(i))

to denote they are for treatments v and v' for the survival rates between sites i and $i + 1$. For example, for $\hat{S}_1$, corr($\hat{\phi}_{1,1}$, $\hat{\phi}_{4,1}$) = 0.6496. In Table 3.9, this correlation is denoted Corr(1, 4, Phi(1)).

Throughout this monograph we emphasize the importance of using the correct model. From the summary results of TEST 1 in Table 3.8, it is clear one should select model H_{2p}; TEST 1.*T*2 rejects model $H_{1\phi}$ as inadequate whereas the sum of TEST 1 components 1.*R*2, 1.*T*3, 1.*T*4, and 1.*T*5 do not reject, thereby corroborating model H_{2p} as a suitable choice. Tables 3.10 and 3.11 show the results under models $H_{1\phi}$ and $H_{2\phi}$, both of which are inappropriate here (keep in mind that, with real data, one does not know what the true model is). The estimators of S_1, S_2, and S_3 will be biased under model $H_{1\phi}$ because this model assumes no treatment effect on the capture probabilities p_{v2}. From Tables 3.9 and 3.10 we extract the following results regarding $\hat{S}$.

	True	model $H_{1\phi}$		model H_{2p}	
Parameter	value	$\hat{S}$	$\hat{\text{se}}(\hat{S})$	$\hat{S}$	$\hat{\text{se}}(\hat{S})$
S_1	0.90	0.763	0.0426	0.889	0.0561
S_2	0.75	0.698	0.0407	0.830	0.0542
S_3	0.70	0.623	0.0383	0.731	0.0502

The estimators under the two models differ substantially. Also, under model $H_{1\phi}$, the 95% CIs for S_1 and S_3 do not cover the true values. Note, also, that the estimated standard errors of the estimates are larger under model H_{2p}.

Table 3.10. – Output from program RELEASE for model $H_{1\phi}$ for the simulated data: scheme B, $k = 6$, 4 groups, lot 1.

Maximum Likelihood Estimates under Model H1Phi

Parameter	Estimate	Standard Error	95% Confidence Intervals Lower	Upper
		Estimates for Group 1		
		Treatment Group 1		
Phi(1)	0.836901	0.063990	0.711480	0.962321
		Estimates for Group 2		
		Treatment Group 2		
Phi(1)	0.766348	0.059922	0.648902	0.883795
		Estimates for Group 3		
		Treatment Group 3		
Phi(1)	0.683631	0.055108	0.575620	0.791642
		Estimates for Group 4		
		Control Group		
Phi(1)	1.097216	0.078790	0.942787	1.251645
		Estimates for Pooled Groups		
p(2)	0.120859	0.009753	0.101743	0.139976

Ratio of Survivals between Groups

Parameter	Estimate	Standard Error	95% Confidence Intervals Lower	Upper
S(1,2,Phi(1))	1.092063	0.069769	0.955316	1.228811
Corr(1,2,Phi(1))		0.658898		
S(1,3,Phi(1))	1.224199	0.081808	1.063855	1.384543
Corr(1,3,Phi(1))		0.639127		
S(1,4,Phi(1))	0.762749	0.042634	0.679187	0.846312
Corr(1,4,Phi(1))		0.717458		
S(2,3,Phi(1))	1.120996	0.077124	0.969834	1.272159
Corr(2,3,Phi(1))		0.624982		
S(2,4,Phi(1))	0.698448	0.040678	0.618720	0.778176
Corr(2,4,Phi(1))		0.701579		
S(3,4,Phi(1))	0.623060	0.038287	0.548018	0.698102
Corr(3,4,Phi(1))		0.680527		

Table 3.11. – Output from program RELEASE for model H_{2p} for the simulated data: scheme B, $k = 6$, 4 groups, lot 1.

Maximum Likelihood Estimates under Model H2Phi

Parameter	Estimate	Standard Error	95% Confidence Intervals Lower	Upper
		Estimates for Group 1		
		Treatment Group 1		
Phi(1)	0.864000	0.127135	0.614816	1.113184
p(2)	0.097222	0.017457	0.063006	0.131439
		Estimates for Group 2		
		Treatment Group 2		
Phi(1)	0.721846	0.107220	0.511694	0.931998
p(2)	0.095588	0.017828	0.060645	0.130531
		Estimates for Group 3		
		Treatment Group 3		
Phi(1)	0.812842	0.151344	0.516208	1.109476
p(2)	0.081197	0.017856	0.046200	0.116193
		Estimates for Group 4		
		Control Group		
Phi(1)	0.989839	0.092805	0.807942	1.171736
p(2)	0.191950	0.021914	0.149000	0.234901
		Ratio of Survivals between Groups		

Parameter	Estimate	Standard Error	95% Confidence Intervals Lower	Upper
S(1,2,Phi(1))	1.196931	0.250256	0.706429	1.687433
Corr(1,2,Phi(1))		0.000000		
S(1,3,Phi(1))	1.062937	0.252253	0.568522	1.557352
Corr(1,3,Phi(1))		0.000000		
S(1,4,Phi(1))	0.872869	0.152296	0.574368	1.171371
Corr(1,4,Phi(1))		0.000000		
S(2,3,Phi(1))	0.888052	0.211517	0.473479	1.302625
Corr(2,3,Phi(1))		0.000000		
S(2,4,Phi(1))	0.729256	0.128095	0.478190	0.980322
Corr(2,4,Phi(1))		0.000000		
S(3,4,Phi(1))	0.821186	0.171188	0.485657	1.156715
Corr(3,4,Phi(1))		0.000000		

Model $H_{2\phi}$ allows unbiased estimation of S_1, S_2, and S_3. However, because model $H_{2\phi}$ is more general than is needed here, the $\hat{S}_1$, $\hat{S}_2$, and $\hat{S}_3$ under that model have larger sampling variances than they have under model H_{2p}. If this loss in efficiency were slight, the best strategy under scheme B would be to use model $H_{2\phi}$. However, from Tables 3.9 and 3.11, we extract the following results regarding $\hat{S}$:

	True	model H_{2p}		model $H_{2\phi}$	
Parameter	value	$\hat{S}$	$\hat{\text{se}}(\hat{S})$	$\hat{S}$	$\hat{\text{se}}(\hat{S})$
S_1	0.90	0.889	0.0561	0.873	0.1523
S_2	0.75	0.830	0.0542	0.729	0.1281
S_3	0.70	0.731	0.0502	0.821	0.1712

The main point of the above is that the standard errors of the $\hat{S}$ are approximately tripled under model $H_{2\phi}$ as compared to the results under H_{2p}. This loss of efficiency is dramatic; one does not want to select model $H_{2\phi}$ under scheme B unless absolutely necessary.