CHAPTER 18

Nest survival models

Jay Rotella, *Montana State University*

In this chapter, we will introduce how to analyze nest survival data with program **MARK**. Nest survival is a key vital rate in the population dynamics of many birds. Accordingly, estimation of nest survival is a key aspect of many studies of breeding bird populations. Given the strong interest in nest survival, there is a rich literature detailing field techniques and estimation methods for this vital rate. The development of estimation techniques has been very active in recent years (e.g., see Dinsmore *et al.* 2002, Rotella *et al.* 2004, and references therein) and so most of the analyses described in this chapter are still relatively new.

At this point, you may be wondering why we need to go to the trouble of using program **MARK** for analyzing nest survival data. After all, with nests you know how many nests you found and you know the fate of every nest. Why not just compare the proportion of successful nests among groups with different attributes? Well, for starters (and many of you probably already know this), it turns out that this is not a valid approach for most data sets. As Harold Mayfield pointed out several decades ago (Mayfield 1961, 1975), such an analysis is only valid if destroyed nests can be found with the same probability as active ones. In most studies successful and unsuccessful nests are not found with equal probability, and most nests are found after egg laying has commenced. Mayfield pointed out that under these circumstances the proportion of successful nests, which he termed apparent nesting success, is biased high relative to actual nesting success, the proportion of nests that survive from initiation to completion. Klett *et al.* (1986:10) provide an excellent illustration of the source of this bias. In fact, it's so good that we'll repeat the essence of the example here.

Imagine a study in which only active nests can be found. Nests are found at various ages (i.e., days since the nest was started). Young leave the nest 35 days after nest initiation (in this example, immediately after hatching, as is typical for species with precocial young). Assume that the probability that a nest survives a single day (daily survival rate; DSR) is 0.96 regardless of calendar date or nest age, and thus, the true probability of a nest surviving from initiation to completion is 0.2396 (0.96^{35}). The following table (at the top of the next page) shows the bias involved in working with the proportion of successful nests if all nests aren't found on or before the day they're initiated. As you can see, overestimation is likely and can be severe.

age of nest when found	probability of surviving to hatching	bias relative to actual hatching success
0	$0.96^{35} = 0.24$	0.00

5	$0.96^{30} = 0.29$	+0.05
10	$0.96^{25} = 0.36$	+0.12
15	$0.96^{20} = 0.44$	+0.20
20	$0.96^{15} = 0.54$	+0.30
25	$0.96^{10} = 0.66$	+0.42
30	$0.96^{5} = 0.82$	+0.58
34	$0.96^{1} = 0.96$	+0.72

Mayfield developed an *ad hoc* estimator of nesting success that overcomes the bias associated with estimates of apparent nesting success. His approach calculates the daily survival probability for only the days that nests were under observation. Thus, the Mayfield model accounts for the fact that some nests are not under observation starting with the day of nest initiation.

To see how Mayfield's method works, consider the following example where 10 nests were found at an age of 0 days, another 10 were found at an age of 14 days, and another 10 were found at an age of 28 days. After being discovered, each nest's fate was checked every 7 days until it failed or reached an age of 35 days. The dataset can be summarized in the following table, which tabulates the number of nests in the sample that survived to a given age – for example, we see that among the 10 nests found at age 14 days, 8 survived to 21 days (i.e., an additional week), 7 survived to 28 days (i.e., an additional 2 weeks), and so on:

initial sample size	age of nest when found	7 days	14 days	21 days	28 days	35 days
10	0 days	7	6	5	3	3
10	14 days			8	7	4
10	28 days					8

The next step is to run Mayfield's calculations on the data. Key steps are to calculate (1) how many nests were unsuccessful and (2) how many days nests were exposed to potential nest failure (termed exposure days). The number of unsuccessful nests is easy and intuitive to calculate: $(10 − 3) + (10 − 4) + (10 − 8) = 15$ failed nests. Exposure days may be a bit less obvious, but here's how it's calculated. All nests in this example were checked every 7 days. So, if a nest survived the interval between 2 visits, it was exposed to potential nest failure on each of the 7 days between visits. Thus, each time a nest survived a re-visit interval, we simply need to add 7 days to the total number of exposure days. For nests found at an age of 0 days, nests survived a total of 24 7-day intervals $(7 + 6 + 5 + 3 + 3)$, which yields 168 exposure days. For nests first found at an age of 14 days, nests survived a total of 19 7-day intervals $(8 + 7 + 4)$, which yields 133 exposure days. For nests first found at an age of 28 days, nests survived a total of 8 7-day intervals, which yields 56 exposure days.

But, what to do with nests that failed between two nest visits? We don't know when the failure occurred and so don't know how many days of exposure to use. Mayfield calculations typically use 1/2 the interval length in the cases of failures that occurred on an unknown date between two nest visits. So, these calculations are pretty simple to make as well. We know that 15 nests failed and, because all intervals in this simple example were 7 days long, each failed nest is assumed to have survived through 3.5 exposure days (one half of the 7-day interval). Thus, we need to add 52.5 exposure days (15 nests x 3.5 days per nest) to the total. Using the values we've just calculated, we see that 15 nests failed during a total of 409.5 exposure days (total = 168 + 133 + 56 + 52.5). Thus, the Mayfield estimate of daily mortality

probability for nests is (15/409.5), or 0.0366. The Mayfield estimate of daily survival probability is simply one minus the daily mortality rate, or (1 – 15/409.5) or (1 – 0.0366) or 0.963. Mayfield's *ad hoc* estimator thus focuses on daily survival probabilities and circumvents the problems inherent in using apparent nest survival probabilities.

Subsequent to Mayfield's (1961, 1975) publications, a variety of authors published a maximum-likelihood approach to estimating daily survival probabilities and nesting success for data from nests that were visited periodically (Johnson 1979, Hensler and Nichols 1981, Bart and Robson 1982). The method is based in statistical theory and provides estimates of the mean and variance of daily survival rate. Thus, use of the maximum-likelihood estimator (MLE) of daily survival probability and its variance is recommended over use of Mayfield's *ad hoc* estimator despite the fact that they produce very similar point estimates of daily survival (and are in fact the same estimator in the case where nests are visited every day).

Let's review some key points made in this chapter so far: we can't validly use the proportion of successful nests as a surrogate for true nest success so we want to use maximum likelihood estimates of DSR and its associated variance instead. As we've seen in earlier chapters, program **MARK** is a very useful tool for doing just this sort of thing. Once we have estimates of DSR, we can raise it to the appropriate power (number of days it takes to go from nest initiation to nest completion [e.g., 35 days]) to estimate true nest success, e.g., $0.96^{35} = 0.2396$. With estimates of the variance of DSR, we can also estimate the variance of nest success ($\text{var}(0.96^{35}) = [(35 \times 0.96^{34})^2 \times \text{var}(\text{DSR})]$. So, by this point we hope you're starting to see how program **MARK** might be useful. But, it gets better because program **MARK** allows one to evaluate a rich variety of competing models of nest survival rate and go well beyond what was possible with basic Mayfield estimation.

Next, we use data collected on Mallard (*Anas platyrhynchos*) nests in North Dakota to introduce the analysis of nest-survival date in program **MARK**. This data set is part of a larger data set collected and analyzed by Stephens (2003) and the same data set used by Rotella *et al*. (2004) in their example analysis. It contains information from 565 nests that were monitored on 18 sites during a 90-d nesting season. Nests of various ages were found during periodic nest searches conducted throughout the nesting season. Once a nest was found, it was re-visited every 4 to 6 days to determine its fate (a binary outcome). For each nest, several covariates of interest were measured: (1) the date the nest was found; (2) the nest's initiation date, which provides information about the age of the nest when it was found, its age during each day of the nesting season, and its expected hatch date (35 days after nest initiation, which is when young typically leave the nest in this species); (3) a measure of how much the vegetation around the nest site visually obscured the nest; (4) the proportion of grassland cover on the 10.4-km^2 study site that contained the nest; and (5-7) the habitat type in which the nest was located (3 indicator variables, each coded as 0 or 1, that were used to distinguish among nests found in 4 different habitat types: native grassland, planted nesting cover, wetland vegetation, and roadside right-of-ways).

18.1. Competing models of daily survival rate

Up to this point, we haven't discussed competing models. We've simply focused on the idea that we need to estimate daily survival rate (DSR) and can use program **MARK** to do so. But, typically we're interested in evaluating competing hypotheses about sources of variation in nest survival. DSR might vary spatially due to changes in features such as habitat characteristics and predator communities. DSR might also be hypothesized to vary temporally. For example, one might hypothesize for a given species that nests are more vulnerable later in the nesting season as predators develop a search image for nests and key into nests as food sources. Thus, DSR may be predicted to be relatively high early in the season, to drop off in the mid-season, and to stay low late in the season. One might also be interested in changes in DSR that are associated with changes in nest age. For a songbird that feeds young in the nest, one

might hypothesize that DSR will be lower during the nestling stage, when parents make many foraging trips per day to and from the nest to provide for begging young, than it will be during egg-laying and incubation stages, when fewer cues are provided to predators. In a multi-year study, we might wonder if DSR differs among years.

It seems safe to say that we can think of lots of possible sources of variation in DSR. So, it also seems reasonable to say that we will want to be able build and evaluate competing models of DSR. This may seem readily obvious, but it was only recently that methods were developed for evaluating multiple-regression type models of DSR (Dinsmore *et al.* (2002); and, yes, these methods were first developed in program **MARK**).

Most studies of nest survival have used Mayfield's *ad hoc* estimator of DSR or the maximum likelihood estimator of Johnson (1979) or Bart & Robson (1982), and these methods assume that DSR is the same for all nests on all dates and for all nest ages (nest fates are independent and identically distributed: *iid*) within the sample. To attempt to meet the assumption of homogeneity of fates within a given sample, one can stratify data prior to analysis and create sets for which fates can reasonably be considered *iid*. For example, data can be analyzed separately for different nesting stages, habitats, years, species, or combinations of these. However, such stratification can be limiting and may prevent researchers from exploring the full suite of models that are likely of interest. This is especially difficult when some of the multiple factors that are thought to affect DSR are measured on continuous rather than categorical scales. Although one can certainly discretize continuous variables and convert them to categorical values, this won't always be desirable.

Fortunately, program **MARK** allows us to model DSR as a function of multiple covariates and to compare competing models. But, before we can begin building models, we need to consider the data input for nest-survival data. It turns out that there are a few unique features to this data type.

18.2. Encounter histories format

Minimally, four pieces of information are required for each nest: (1) the day of the nesting season on which the nest was found; (2) the last day the nest was checked when alive; (3) the last day the nest was checked; and (4) the fate of the nest (0 = successful, 1 = depredated). Program **MARK** uses these key pieces of information to generate an encounter history for each nest in live/dead (**LDLD**) format (see Chapter 2). Although you'll never see the **LDLD** type encounter histories in program **MARK** for this data type, we will explore them a bit more below as they can help us understand what's taking place with nest survival data.

The days mentioned above refer to standardized days within the study's nesting season. These standardized days are obtained by first calculating the earliest date on which you began recording data on nest survival, treating this date as day 1 of the nesting season, and then sequentially numbering all subsequent days for which you observed nests up to the last date of data collection. For the Mallard data set used in this example, data collection began on 24 April and continued for 90 days. So, 24 April was standardized as day 1 and the last day was day 90.

The next step is to create an encounter history for each nest based on the critical information described above. In the mallard study, the first nest was found on day 1 (April 24) of the season with 1 egg. This nest was observed periodically over the next 35 days and was successful. Thus, this nest had the following encounter history:

```
/* 1 */ 1 35 35  0 1;
```

Here, the encounter history begins with a comment (/* 1 */) referring to the nest's identification number (this is quite handy when data checking but not essential to provide). The key information

follows the comment: (1) first is the day the nest(s) was/were found, labeled time i; (2) next is the last day the nest(s) was/were known to be present, labeled time j, which for successful nests should be the day the nest attempt was successfully completed (as it is in this example); (3) then comes the last day that the nest(s) was/were checked or, for successful nests, the day that the nest attempt should have been successfully completed (as it is in this example), labeled time k; (4) next is the fate of the nest(s): **0** means successful, and **1** means destroyed or unsuccessful; and (5) the last value is the number (frequency) of nests that had this history (this will usually be 1 but doesn't need to be if multiple nests have the same encounter history).

Later in the study, a nest had this encounter history:

```
/* 1931  */ 63 81 85 1 1;
```

From this, we can see that this nest was found on day 63 of the nesting season, last known to be active on day 81, and last checked on day 85. The nest was unsuccessful and was destroyed sometime between day 81 and day 85.

You may have noticed that for successful nests, all the information about survival time is contained in times i and j. If you did, you're paying attention and you're right. Time k is only used for unsuccessful nests to bracket the interval when the nest was destroyed. Another point worth emphasizing is the fact that dates j and k should not extend beyond the nest's successful completion date even if the actual nest check was done after the hatch date (or, in the case of nidicolous species with a nestling stage, beyond the expected fledging date). If you put in the actual check dates rather than completion dates, you'll give the nest credit for surviving days when it was no longer subject to possible failure – definitely not something we want to do.

The histories above could be extended to include individual covariates after the 5 required variables. In the mallard example, the following covariates were added to the encounter histories: (1) a measure of how much the vegetation around the nest site visually obscured the nest; (2) the proportion of grassland cover on the 10.4 km^2 study site that contained the nest; (3) an indicator variable that was coded as **1** if the nest was in native grassland and **0** otherwise; (4) an indicator variable that was coded as **1** if the nest was in planted nesting cover and **0** otherwise; (5) an indicator variable that was coded as **1** if the nest was in wetland vegetation and **0** otherwise; and (6) the age of the nest on the first day of the nesting season. Because the age variable is for the first day of the nesting season, this value will be negative for all the nests in the sample except those that were initiated (started) on or before day 1.

Several items are noteworthy about the covariates. First, date, which may be of interest as a covariate in some models, is incorporated into the required fields of the encounter history, and so, there is no need to include date information elsewhere. One might, of course, wish to have a variable(s) for year of study in a multiple-year study. Second, in this example, there were 4 habitat types of interest but only 3 indicator variables are included: the 4th habitat type (roadside right-of-ways) is indicated if all 3 of the indicator variables have a value of **0**. Third, the age variable as described above might seem a bit odd at first glance. Why do we want to work with the age of each nest on the first date of the nesting season? The short answer is it allows us to generate each nest's age on every other day of the nesting season and these are critical to have if we think that DSR might vary by nest age. A longer answer comes below when we use the age variable in modeling.

Is that it for covariates? Well, no. As we've seen for some other data types, the encounter histories are assigned to groups, so we could also group the data in various ways, which is another way of incorporating a group covariate. Group covariates can also be incorporated through the design matrix in ways that will be explained below.

In the mallard example, data were originally recorded in interval-specific form, i.e., each row of data contained information for one observation interval for an individual nest. For the 2nd nest found in the

study, the following rows of raw data were recorded (shown at the top of the next page).

Nest ID	Species	Study Site	Habitat Code	Obs	t	IFate	SDate	Sage	Robel	PpnGR
2	MALL	14	PICov	1	5	0	1	3	0.88	0.96
2	MALL	14	PICov	2	5	0	6	8	0.88	0.96
2	MALL	14	PICOv	3	4	0	11	13	0.88	0.96
2	MALL	14	PICov	4	6	1	15	17	0.88	0.96

Here the columns represent (1) the nest's identification number, (2) a species code, (3) the study site, (4) a habitat code, (5) a number indicating which observation interval is represented on this row for this nest, (6) the number of days in the observation interval, (7) the nest's fate for this observation interval (0 = successful, 1 = failed), (8) the day of the nesting season at the start of the interval, (9) the age of the nest at the start of the interval, (10) a measure of how much the vegetation around the nest site visually obscured the nest, and (11) the proportion of grassland cover on the 10.4 km^2 study site that contained the nest.

The data for nest #2 were re-formatted for analyses in **MARK** as:

```
/* 2 */ 1 15 21 1 1 0.88 0.96 0 1 0 3;
```

To create the actual input file for **MARK**, the data for each nest to be used in the analyses must be formatted appropriately. In addition, the first non-comment line of the file needs to consist of the statement 'Nest Survival Group=1;'. Following this command are the data for all of the nests in the 1st group. When this is done for the mallard data, the first portion of the encounter history file appears as follows:

```
Nest Survival Group=1;

/* 1 */  1  35  35  0  1  4.500  0.9600  0  1  0  1;
/* 2 */  1  15  21  1  1  0.875  0.9616  0  1  0  3;
/* 4 */  1  11  15  1  1  2.750  0.9616  0  1  0  3;
/* 5 */  1  31  33  1  1  3.375  0.9616  0  1  0  3;
```

Note: The **MARK** help file, and Chapter 2 of this book, contains much useful information about how to build the encounter history file. It is definitely worth taking the time to review the information provided.

18.3. Nest survival, encounter histories, & cell probabilities

In **MARK**, the data type for the encounter history is LDLD (although you won't code it this way in the INP file - **MARK** will handle this for you). The cell probability is modeled as the product of the survival probabilities from time (i) to time ($j - 1$). For successful nests, i.e., fate = 0, the last time seen alive (j) and the last time (k) should always be the same and equal to the day the nest attempt was completed (which may be hatch date for species with precocial young, or fledging date for species with altricial young). For unsuccessful nests, fate > 0, $j < k$.

For successful nests, time (k) is ignored. For situations where the date of hatch is not known exactly, none of the times should exceed the potential nest completion date, because the nest should not be

considered at risk of failure when a nesting attempt has already been completed. For unsuccessful nests, the cell probability for the nest is taken as the product of the survival rates from time (i) to time ($j - 1$), times 1 minus the product of survival from (j) to ($k - 1$).

The following series of examples for 5 occasions will clarify how to code nest survival data with the triplet i, j, and k.

triplet	*history and interpretation*
$i = 1, j = 3, k = 5, fate = 1$	1010100001, with cell probability $S_1 S_2 (1 - S_3 S_4)$
$i = 1, j = 3, k = 3, fate = 0$	1010100000, with cell probability $S_1 S_2$
$i = 1, j = 3, k = 3, fate = 1$	invalid, because the nest was observed both present and destroyed on day 3
$i = 1, j = 1, k = 3, fate = 1$	1000010000, with cell probability $1 - S_1 S_2$
$i = 1, j = 1, k = 3, fate = 0$	invalid because the nest was observed present only on day $i = 1$. If the nest was still present on day 3, then j should have been coded as $j = 3$
$i = 1, j = 3, k = 5, fate = 0$	partially invalid because the nest was observed for the interval 1 to 5, when the nest was successful, but the coding shown will only use the data from 1 to 3, giving 1000100000, with cell probability $S_1 S_2$. For $fate = 0, j = k$
$i = 3, j = 3, k = 3, fate = 0 \ or \ 1$	invalid because the nest was not observed over an interval for either fate

The key difference between *known-fate* (discussed in the preceding chapter) and *nest survival* data types is that with nest survival data, we don't know exactly what day the nest was destroyed for cases where the nest was unsuccessful. Thus, consider the cell probability for the following:

```
/*GG00, 1995-076*/    53  59  63  1  14;
```

This nest was found on day $i = 53$, checked and found still present on day $j = 59$, and found destroyed (i.e., failed) on day $k = 63$. The last variable (an individual covariate) in this example is the age of the nest at the time it was first found – 14 days old for the first nest.

The cell probability corresponding to this nest would look like:

$$S_{53}S_{54}S_{55}S_{56}S_{57}S_{58} \left(1 - S_{59}S_{60}S_{61}S_{62}\right)$$

The first portion of the expression models the survival of the nest from day $i = 53$ to day $j = 59$. The second portion of the expression, in brackets, models the failure of the nest during the interval from day $j = 59$ to day $k = 63$. That is, had the nest been successful during this interval, the probability would have been $S_{59}S_{60}S_{61}S_{62}$. Because the nest was destroyed at some time during that interval, this quantity is subtracted from 1. The complete encounter history for this example would consist of 52 pairs of '00', followed by the following string with blanks inserted to delimit occasions denoted on the second line:

```
10 00 00 00 00 00 10 00 00 01
53 54 55 56 57 58 59 60 61 62
```

Needless to say, having to write this out in full on your own would be a rather tedious exercise. Fortunately, **MARK** handles all of this for you 'behind the scenes'. But, it is important to understand how the probability statements are created.

———————————————————— begin sidebar ————————————————————

Effective sample size and nest survival analyses

The effective sample size for AIC$_c$ is computed somewhat differently for the nest survival model than other models in **MARK**. Typically, each binomial trial in a model provides one degree of freedom. For the nest survival model, each day that the nest is known to survive contributes 1 degree of freedom because each day is a binomial trial where the result is known. However, the interval in which a nest fails only contributes 1 degree of freedom also, because the exact day of failure is not known, but only that the nest failed during the interval.

Thus, a record like

```
/*YGBB, 1995-074*/          44  55  55  0    1  18;
```
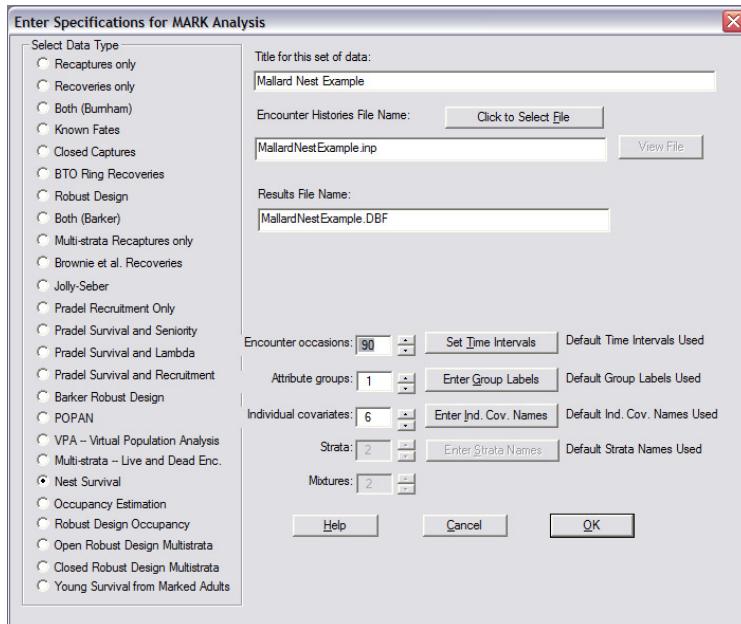
contributes 10 degrees of freedom, because there are 10 binomial trials where the nest was known to succeed. However, the following record only contributes a single degree of freedom, because the nest failed somewhere in the interval 58-61.

```
/*WGGW, 1995-078*/       58  58  61  1    1  18;
```
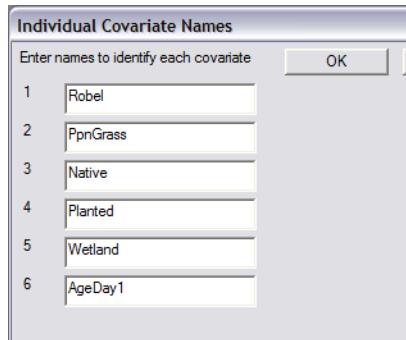
———————————————————— end sidebar ————————————————————

18.4. Building models

Here we introduce the steps for fitting nest survival models in **MARK**. First, we create a new **MARK** database file from your nest-survival encounter history file using the '**File | New**' menu option. Choose the '**Nest Survival**' data type and enter the required information. For our mallard example, we have a 90-occasion study (data were collected across 90 days), 1 attribute group, and 6 individual covariates.

If you'd like, you can enter names for the individual covariates. As we have 6 that we'll use in the model-building exercises below, it's probably best to name them. So, choose to 'Enter Individual Covariate Names' and enter the following: Robel, PpnGrass, Native, Planted, Wetland, and AgeDay1.

Model 1: Constant Daily Survival Rate

Now that we have created a new database, we'll start our model building with the simplest model, the maximum likelihood version of the Mayfield model. Remember, in this model, we assume that all nests in the sample under consideration have the same DSR on every day. So, we simply need to constrain DSR accordingly.

This can be done with the PIM chart, the survival PIM window, or by choosing to run the appropriate pre-defined model, which is listed appropriately as the S(.) model (although, again – as noted before, we do not favor relying on pre-defined models – they're available in **MARK** to expedite analysis of common models, and should only be used if you have a firm grasp on what you're doing).
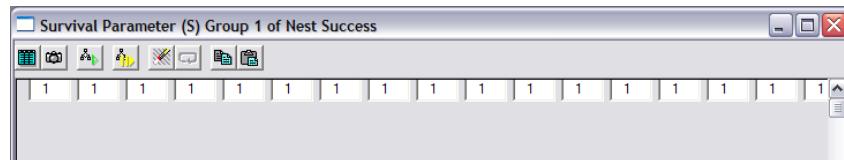
───────────────────── begin sidebar ─────────────────────

PIMs and nest survival

This brings up a question: how many PIMs are there for this nest survival dataset? Well, it turns out that there is only one. That's right: we've only got one group, and there are no parameters other than survival (DSR) to estimate for this data type. It also brings up the question of how many cells there are in the PIM. For this dataset, there are 90 encounter occasions and thus, there are 89 cells in the PIM to estimate the survival probabilities across the 89 intervals. Unfortunately, for this example, the PIM is so wide we can't show you the whole thing – but, try it yourself.
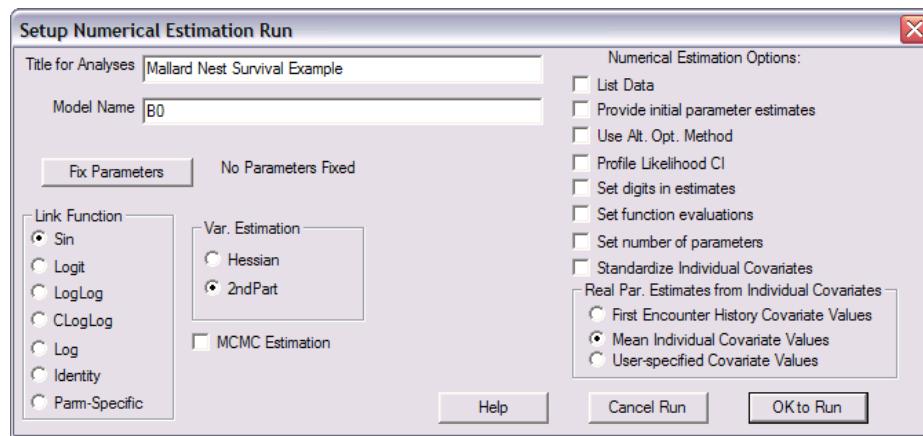
───────────────────── end sidebar ─────────────────────

OK, back to model building. To set up this model using the PIM window, simply set all the values in the PIM to 1, which, of course, is the $S(.)$ model we're after here. The easy way to set all the values to 1 is to right-click the PIM window, and select '**Constant**'.

Next, go ahead and run this model. You might name the model something like 'Constant DSR' or

'B0' (as it is an intercept-only model, and we usually denote the intercept as β_0 or β_1 (depending on your convention), or 'B0' or 'B1'in ASCII, respectively. Here we will use β_0 for the intercept.). Be sure to turn off the option '**Standardize Individual Covariates**', because this option is not needed for these data (i.e., the covariate values are scaled to not produce numerical problems), and because we will not want to scale the AgeDay1 variable in a future analysis. So, to maintain compatibility of the models, we uncheck the '**Standardize Individual Covariates**' box.
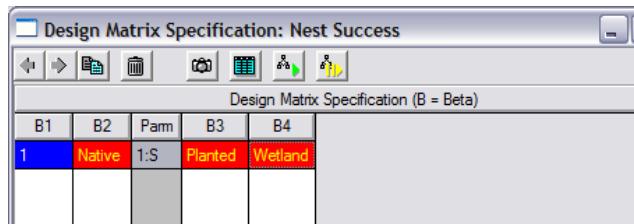
Once the model has run, you can examine the real parameter estimates whereupon you'll see the following:

This is the estimate of DSR when DSR is constrained to be constant across all nests and all dates in the sample. For mallards, it is typically considered that a nest must survive 35 days in order to make it from the 1st day of egg laying to nest completion, which is the day the young leave the nest. So, you could raise the estimated DSR to the 35th power to obtain a point estimate of nest success, which is 0.184.

Model 2: DSR Varies by Habitat Type

Often there is interest in whether or not DSR varies among nests found in different habitat settings. Although there are certainly many metrics that can be used to describe the habitat conditions associated with a nest, researchers often categorize habitat conditions and use a metric such as habitat type. In the study of mallard nests, the researchers used 4 habitat types, which you'll remember were described using 3 dummy variables as individual covariates. To build a model that allows DSR to vary by habitat type, you'll need to use the design matrix and the individual covariates that tell **MARK** which habitat type each nest was in. The following design matrix (top of the next page) can be used.
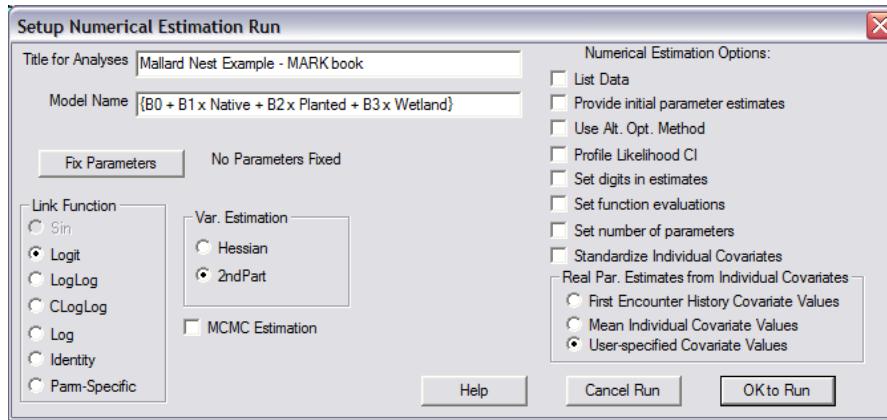
DM, groups and nest survival
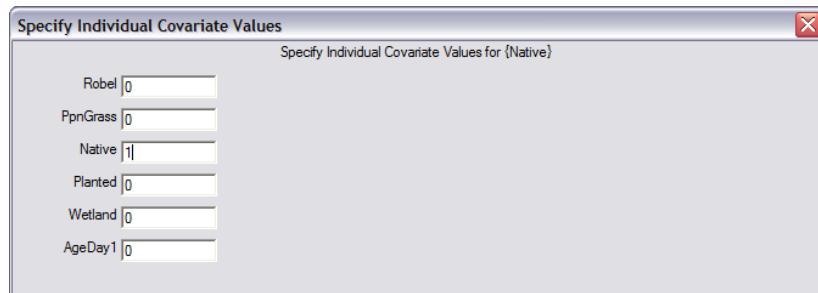
If you have experience using groups in **MARK**, you may recognize that encounter histories for nests in different habitats could be put in different groups. If this were done, you would have a different PIM for DSR for each group and could evaluate a model that allows DSR to vary among habitat types without using a design matrix. In this example, all of the data were entered in a single group, the habitat type associated with each nest is input as an individual covariate, and the design matrix is used to build models in which DSR is allowed to vary among habitat types. The modeling results obtained will be the same regardless of whether habitat type is input using different groups or individual covariates.

Once you've specified the design matrix appropriately, you're almost ready to run this model. When you name the model, you might call it 'Habitat Type' or, in keeping with the naming strategy used in the previous model, you might call it 'B0 + B1 x Native + B2 x Planted + B3 x Wetland'. Once again, be sure to turn off the option '**Standardize Individual Covariates**', because this option is not needed for these data. Next, you need to think through which habitat you'd like the estimates of the real parameters to be for. One easy way to control this is by specifying that you want to use '**User-Specified Covariate Values**' (check the button in the lower right corner of the setup window. For more on this topic, look in the **MARK** help file under '**Individual Covariates and Real Parameter Estimates**.').

Let's assume for now that you want to get estimates for Native vegetation (we'll get estimates for other habitat types later). Specify that you want to estimate the real parameters when Native = 1, i.e., when the habitat type is native grassland. For this model, all other parameter values can be left as 0: values for Robel, PpnGrass, and AgeDay1 are ignored as they're not in the model; values for Planted and Wetland need to be 0 to ensure we're only estimating DSR for nests in Native.
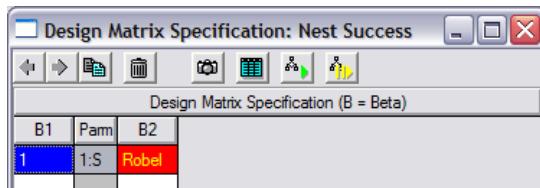
We want to know the habitat-specific estimates of DSR. To get those, we'll need to run the model 3 more times and choose to base the real parameter estimates from user-specified covariate values. When you run this model, you'll see that (1) the results for this model suggest that the simpler intercept-only model (our 'S(.)' or 'B0' model) is more appropriate and (2) the estimate of DSR for native grassland is 0.946 (SE = 0.005). Now, you can re-run the model 3 more times: setting 'Planted=1' and all other variables to 0 on one run, setting 'Wetland=1' and all other variables to 0 on another run, and then finally running the model with all variables set to 0. What will this last model run allow us to estimate? If you figured out that this will allow you to obtain an estimate for roadside habitat (i.e., the habitat that a nest is in for this example if it's not in planted cover, native cover, or a wetland), then you're getting this (or this is old hat to you and you might consider reading something else!).

Notice that you don't want multiple versions of this same model sitting in the results browser as they influence the model weights. But, you may want to look at each and see what the habitat-specific estimates of DSR are (Native: 0.946 [SE = 0.005]; Planted: 0.956 [SE = 0.003]; Wetland: 0.951 [SE = 0.011]; and Roadside: 0.956 [SE = 0.009]).

You now see how **MARK** can be used to evaluate basic models of DSR using maximum likelihood estimation and AIC. This alone would make **MARK** a very useful tool for analyzing nest survival data. However, we can do much more. To showcase a few of the other types of models that can be run, we'll now explore models that allow DSR to vary based on factors such as habitat metrics measured on a continuous scale.

Model 3: DSR varies with vegetation thickness (continuous covariate)

The input file contains an individual covariate labeled Robel (a continuous measure of how much the vegetation around the nest site visually obscured the nest). One way to evaluate whether DSR varies with Robel is to use the following design matrix:


When this model is run (you might call it 'B0 + B1 x Robel'), you find that this model receives less support than the simpler intercept-only model (our S(.) model).

Model 4: DSR varies with the amount of native vegetation in the surrounding area

The input file contains an individual covariate labeled `PpnGrass`, which you may recall is a continuous measure of the proportion of grassland cover on the 10.4 km^2 study site that contained the nest. Given interest in relationships between nest survival and spatial features of vegetation, models containing such covariates may be of interest in some studies. And, as you can probably now readily envision, such models are readily evaluated in **MARK**.

You can use the design matrix to build a model containing `PpnGrass` and if you do, you will find that this model receives more support than any of the models discussed so far. In fact, there is evidence that DSR is higher on sites that contain more grassland cover than on sites with less cover.


```
mrk6353z.tmp - Notepad
File  Edit  Format  View  Help
                    Mallard Nest Example - MARK book

       LOGIT Link Function Parameters of {B0 + B1 x PpnGrass}
                                             95% Confidence Interval
 Parameter              Beta        Standard Error      Lower           Upper
 -----------------   -------------  --------------   --------------   --------------
     1:              3.0123817       0.0580504        2.8986028        3.1261605
     2:              0.1082289       0.0557178       -0.9779973E-03    0.2174359
```

But, we'll assume that we have other *a priori* models that are of interest and will run them all before getting into the details about inferences that can be drawn from the analysis.

Model 5: DSR varies across the nesting season

One might predict that nests will be more or less vulnerable as the season progresses. Changes might be predicted to occur because the abundances of predators and alternate prey are thought to vary seasonally or because vegetation characteristics may change dramatically over the course of the season in some habitat settings. Certainly other interesting hypotheses might exist as well. For now, let's just evaluate a model that constrains DSR across the nesting season such that it is forced to follow a trend (a linear trend on the logit scale).

To build a trend model, go ahead and open the PIM window and set the values to all different (i.e., so that the values in the 89 cells of the PIM take on the values 1, 2, 3, . . . , 89). Next, open a reduced design matrix with 2 columns. Now, because our PIM had 89 parameters, the design matrix has 89 rows and 2 columns. Fill the first column in with 1's (we want to estimate a β term for our intercept) and the second column in with numbers running from 1 to 89 (we want to estimate a β term for a slope indicating how the log-odds of DSR change with increasing season date; log-odds are discussed in Chapter 6).

Go ahead and run this model (you might call it '`B0 + B1 x Date`'). You can see that there's not much support for this model and that the real estimates don't indicate any large changes in DSR over the season. Even though in this case we don't see much seasonal variation in DSR, you can probably envision how this model might be of interest in other studies. Also, you might also be interested in pursuing curvilinear versions of this model (e.g., a model with a quadratic term) where the middle of the nesting season is expected to have the highest (or perhaps the lowest) DSR. If you did want to evaluate a quadratic term you would create a third column in the design matrix and fill it with the squared values of 1, 2, 3, . . . , 89, which, of course, are 1, 4, 9, . . . , 7921.[*]

---

[*] An easy way to do this is to build an Excel spreadsheet that has columns containing the values and to copy and paste them into the design matrix in **MARK**.

Chapter 18. Nest survival models

Model 6: DSR varies with nest age

One might predict that nests will be more or less vulnerable as they age. This could be of interest because the species is known to provide predators with fewer and fewer cues each day as the nest ages. Perhaps this occurs because the species is known to increase nest attentiveness and to take fewer breaks from incubation as hatching nears. In contrast, in other species, one might predict that DSR is lower later in nesting when nestlings and feeding of nestlings may provide more cues to predators. Obviously, the predictions will vary by species. Regardless, we can use **MARK** to model a variety of models that allow DSR to vary with nest age.

Let's examine a simple model that allows DSR to follow a trend in accordance with nest age. To build this model, we first need to let DSR be different on each day of the nesting season. To do this, we once again need a PIM that is numbered from 1 to 89. Next, open a design matrix with 2 columns. Fill the first column of the design matrix with 1's (our intercept). Now for something that is new, at least in this chapter. We want to build a model that contains each nest's age on each day of the nesting season. Fortunately, we have the covariate `AgeDay1`, which you may recall tells us the age of the nest on the 1st day of the nesting season (a negative value for most of the nests!). We will now use this covariate and one of **MARK**'s handy design matrix functions to fill in our design matrix.

In the **MARK** help file (or in Section 11.4 of Chapter 11 of this book), you should read over the section on *design matrix functions* – you can find it by searching for '**Design Matrix Add and Product Functions**' in the helpfile index. We will use the '**add**' function, which not surprisingly is a function that adds 2 arguments together.

To create an individual covariate that is each nest's age on each day of the nesting season, we will start with `AgeDay1` and simply add 0 to it on the first day of the nesting season, 1 to it on the 2nd day of the nesting season, 2 to it on the 3rd day of the nesting season,. . ., and 88 to it on the 89th day of the nesting season. The actual function takes the form '**add(argument1,argument2)**' where the 2 arguments in this case are `AgeDay1` and a continuous string of discrete numbers from 0 to 88. Some of you may be concerned about the fact that negative values were entered for most nest ages on day 1 of the nesting season. But, the negative values are never actually used to compute a survival, because a nest doesn't enter the likelihood until the '**add(AgeDay1,x)**' value is zero, assuming that the correct age on day 1 (i.e., the correct negative value) has been entered. This business of adding to a negative age may be conceptually difficult at first but it does work!

The completed design matrix appears as follows:

Go ahead and run this model (you might call it '`B0 + B1 x Nest Age`'). Be sure to uncheck the '**Standardize Individual Covariates**' box, because otherwise the adding of the values in each row of the design matrix will not perform as you are expecting. You'll see that the results indicate that DSR might vary with nest age. But first, we might just want to run a few models that consider multiple

covariates at once.

Note that you can also construct a model with age in a quadratic (or higher power function). Simply add a third column to the design matrix, and enter '`power(add(AgeDay1,0),2)`' in the 1st row of the 3rd column, '`power(add(AgeDay1,1),2)`' in the 2nd row of the 3rd column, and so on, ending with '`power(add(AgeDay1,88),2)`' in the 89th row of the 3rd column. (As noted above, in both Chapter 11 of this book and in **MARK**'s help file for design matrix functions, there are instructions for how to create such a column efficiently in Excel using Excel's '`concatenate`' function).

Model 7: models with multiple covariates

Let's imagine that the researchers had predicted *a priori* that DSR might vary with nest age and each of the various habitat measures (`Habitat Type`, `Robel`, and `PpnGrass`). So, let's build 3 more models by simply adding to the design matrix used for the '`B0 + B1 x Nest Age`' model. We can do this using the '`Add Column`' function (or the combination of the `ctrl` key and the '`A`' key) while the design matrix window is active.

Let's go ahead and run these 3 models and call them: (1) '`B0 + B1 x Nest Age + B2 x Robel`', (2) '`B0 + B1 x Nest Age + B2 x PpnGrass`', and (3) '`B0 + B1 x Nest Age + B2 x Native + B3 x Planted + B4 x Wetland`'.

Model	AICc	Delta AICc	AICc Weight	Model Likelihood	No. Par.	Deviance
{B0 + B1 x Nest Age + B2 x PpnGrass}	1563.0095	0.0000	0.46477	1.0000	3	1557.0056
{B0 + B1 x Nest Age}	1564.0660	1.0565	0.27405	0.5896	2	1560.0640
{B0 + B1 x Nest Age + B2 x Robel}	1565.9055	2.8960	0.10924	0.2350	3	1559.9016
{B0 + B1 x Nest Age + B2 x Native + B3 x Planted + B4 x Wetland}	1567.3435	4.3340	0.05323	0.1145	5	1557.3337
{B0 + B1 x PpnGrass}	1567.3678	4.3583	0.05258	0.1131	2	1563.3658
{B0}	1569.1166	6.1071	0.02193	0.0472	1	1567.1159
{B0 + B1 x Robel}	1570.7749	7.7654	0.00957	0.0206	2	1566.7729
{B0 + B1 x Date}	1570.8265	7.8170	0.00933	0.0201	2	1566.8245
{B0 + B1 x Native + B2 x Planted + B3 x Wetland}	1571.9571	8.9476	0.00530	0.0114	4	1563.9506

It seems pretty clear that it's worth considering nest age and that `PpnGrass` appears to be important as well. We'll look at the details of model results a bit more below, but first, let's consider another factor possibly affecting DSR that one might want to investigate.

18.4.1. Models that consider observer effects on DSR

Although we won't go into all the details here, it seems worth pointing out that **MARK** also allows one to readily evaluate possible effects of observer visits to nests on DSR if the appropriate information is entered in the encounter history. In the case of the mallard study, 89 individual covariates could be added to the encounter history that simply are coded as 0 or 1 and that indicate whether a nest was visited on each day of the nesting season or not. These 89 covariates might be called something like `VisitDay1`, `VisitDay2,...`, `VisitDay89`. The idea being that we could then build a design matrix that modeled DSR as a function of whether or not a nest was visited on a given day.

For example, if these covariates were included in the dataset, you could build a design matrix like the one shown at the top of the next page.
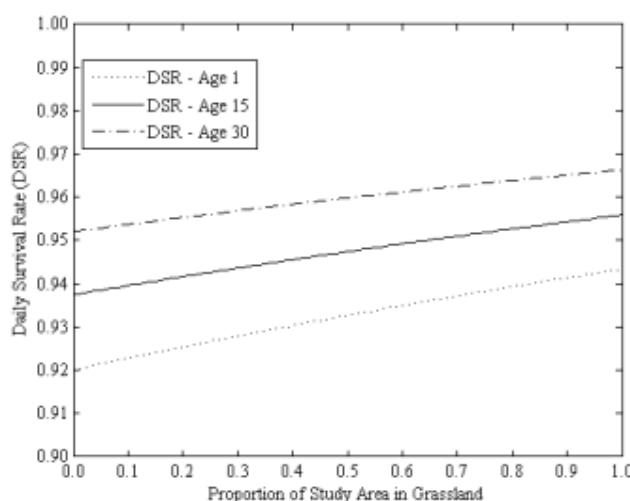
Further, you could include consideration of nest visits at the same time that other covariates were also considered. In cases where there is concern that nest visits may affect DSR on the day of the visit, such models may be of considerable interest. Other versions of such models could be built to consider more complex effects of visits on DSR, e.g., effects that are strongest immediately after a visit but that persist for multiple days.

Also recognize that you definitely do not want to use the '**Standardize Individual Covariates**' option with the model shown in the above design matrix. If you were to standardize the individual covariates, each of the 89 variables `VisitDay1` through `VisitDay89` would be standardized differently, and as a result, the model would be nonsense. That is, the β_i value (β_2) would be multiplying variables with different meanings in each row. So, by not standardizing the individual covariates, the meaning of `VisitDay1` is the same as the meaning of `VisitDay2`, etc.

18.5. Model results

The most parsimonious model of DSR included `Nest Age` and `PpnGr`. This model was 1.06 AIC_c units better than the second-best model, which included `Nest Age` but not `PpnGr`, and was > 2.90 AIC_c units better than all other models evaluated. Models that held daily survival probability constant or simply allowed it to vary by habitat type, i.e., the only model types that have been used in many recent publications on nest survival, received little support ($\Delta AIC_c > 6.11$). The best model indicated that DSR increased with nest age ($\hat{\beta} = 0.0188, \text{SE} = 0.008$) and grassland extent ($\hat{\beta} = 0.369, \text{SE} = 0.211$), as can be seen in the following graphic:
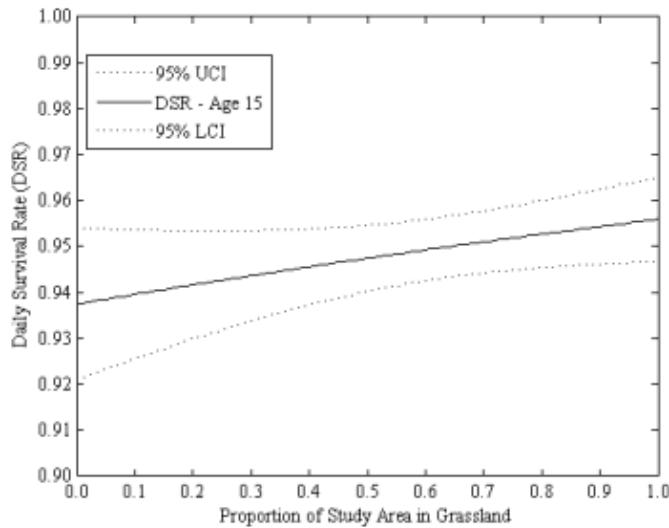
Note: individual covariate values were not standardized in the analysis that produced these estimates so that actual nest ages and levels of `PpnGrass` could be used when producing the estimates; convergence was not problematic here.

When the logit link is used, our estimate of DSR for any combination of nest age and `PpnGrass` is simply

$$\frac{\exp\left(\beta_0 + \beta_1(\texttt{Nest age}) + \beta_2(\texttt{PpnGrass})\right)}{1 + \exp\left(\beta_0 + \beta_1(\texttt{Nest age}) + \beta_2(\texttt{PpnGrass})\right)}$$

So, for example, the estimated DSR for age=15:

So, the estimates presented in the figure can readily be obtained by substituting in the estimates of the β's and various values of `nest age` and `PpnGrass`.

Of course, one would want to see some measure of uncertainty on these graphs as well. So, it is worth considering how we can obtain estimates of the SE associated with various estimates of DSR. This is a bit more complicated because now there are 3 estimated β's and an associated (3×3) variance-covariance matrix for those estimates.

One method of incorporating the variances and covariances associated with the 3 β's is the 'Delta method' (Seber 1982). Although we won't go into the use of the Delta method in this chapter (it is described in detail in Appendix B*), the figure (above) provides estimates of DSR and 95% confidence intervals estimated by the Delta method for 15-days-old nests for different levels of `PpnGrass`.

18.6. Individual covariates and design matrix functions

In chapter 11 (section 11.4) we introduced several special functions that you can use in the design matrix, to make it easier to build specific models. These functions are especially useful for building complex models involving individual covariates. In Chapter 11 (section 11.4), there is a fairly comprehensive review of how to use several of these (specifically, the `add` and `ge` function) for a nest survival analysis

---

* Complete details for the construction of CI for parameter estimates that are functions of covariates can be found in Chapter 6, and Chapter 11.

which has fairly common design considerations.

18.7. Additional applications of the nest success model

Another application of the nest success model is to estimate the survival of radio-tracked animals when the animals are not monitored in discrete intervals, as required by the known-fate data type (see the preceding chapter on *known-fate analysis*). Consider that such data are no different than a set of nests where all the nests are not visited on the same day. A DSR is estimated for each day of the study based on the sample of animals available on that day, and the exact day of death is not required (just as the exact day that a nest was destroyed is not known). We call these kinds of data 'Ragged Telemetry Data' because the sampling scheme is ragged, but useful estimates can still be obtained.

18.8. Goodness of fit and nest survival

Like the known fate data type, the saturated model for the nest survival data type is a useful model. Because the saturated model fits the data perfectly, there is no GOF test provided in Program **MARK**. However, a goodness of fit test for nest survival data was developed recently by Sturdivant *et al.* (2007), and it can be implemented with published **SAS** code, but not in **MARK**. Additional details, including code and sample datasets, can be found at: `http://www.montana.edu/rotella/nestsurv/`.

18.9. Summary

For nest survival data, Program **MARK** provides an excellent alternative to traditional constant-survival methods that have been used in most studies of nest survival to date. The methods shown in this chapter (1) can be used to conduct analyses of stratified data (appropriate if the simplifying assumptions of constant survival apply) and provide estimates that are almost identical to Mayfield estimates (or various refinements), (2) permit comparisons of survival probabilities among groups, (3) allow a much broader variety of covariates and competing models to be evaluated, and (4) should be employed in most nest-survival studies. It is worth noting that these methods can also be used for analyzing survival data collected from radiomarked individuals using ragged (uneven) intervals among animals and over time. The **MARK** help file provides more comments on this topic.

Despite these advances, further analysis improvements would be useful. Improved methods of estimating goodness-of-fit and for detecting and estimating overdispersion, or extra-binomial variation, would be useful given that a variety of factors may cause overdispersion. Nest-success data are commonly collected according to multilevel designs that result in grouped data, e.g., multiple observations on at least some nests, multiple nests per site, and multiple sites within each year. Thus, undetermined random effects of individuals, sites, and years could cause overdispersion or within-group correlations in daily survival probabilities, e.g., nest fates from multiple nests from within a colony or from a given study plot may not be independent. In addition, the spatial clustering of covariate levels could generate spatial correlation in nest survival rates and thus cause overdispersion. The random-effects model described by Rotella *et al.* (2004) can estimate random effects due to one source, e.g., site.

However, even these methods do not accommodate multi-level nonlinear mixed models (e.g., some random effects associated with site, some associated with year, and others associated with individual nests), although, as mentioned above, they will be of interest in some studies (*note*: This is a good example of a situation where the MCMC tool in **MARK** could be used, but further discussion of that complex topic is beyond the scope of this chapter). Finally, in some studies, uncertainty will exist about nest ages

and when transitions among nest stages occur (Williams *et al*. 2002). This problem has been addressed for stratified data (Stanley 2000, 2004) but not yet for data with more complex covariates.

18.10. References

Bart, J., and Robson, D. S. (1982) Estimating survivorship when the subjects are visited periodically. *Ecology*, **63**, 1078-1090.

Dinsmore, S. J., White, G. C., and Knopf, F. L. (2002) Advanced techniques for modeling avian nest survival. *Ecology*, **83**, 3476-3488.

Klett, A. T., Duebbert, H. F., Faanes, C. A., and Higgins, K. F. (1986) Techniques for studying nest success of ducks in upland habitats in the prairie pothole region. United States Fish and Wildlife Service Resource Publication No. 158.

Hensler, G. L., and Nichols, J. D. (1981) The Mayfield method of estimating nest success: A model, estimators and simulation results. *Wilson Bulletin*, **93**, 42-53.

Johnson, D. H. (1979) Estimating nest success: the Mayfield method and an alternative. *Auk*, **96**, 651-661.

Mayfield, H. F. (1961) Nesting success calculated from exposure. *Wilson Bulletin*, **73**, 255-261.

Mayfield, H. F. (1975) Suggestions for calculating nest success. *Wilson Bulletin*, **87**, 456-466.

Rotella, J. J., Dinsmore, S. J., and Shaffer, T. L. (2004) Modeling nest-survival data: a comparison of recently developed methods that can be implemented in **MARK** and **SAS**. *Animal Biodiversity and Conservation*, **27**, 187-204.

Seber, G. A. F. (1982) *The Estimation of Animal Abundance and Related Parameters. 2nd ed*. Macmillan, New York, New York, USA. 654 pp.

Stanley, T. R. (2000) Modeling and estimation of stage-specific daily survival probabilities of nests. *Ecology*, **81**, 2048-2053.

Stanley, T. R. (2004) Estimating stage-specific daily survival probabilities of nests when nest age is unknown. *Auk*, **121**, 134-147.

Stephens, S. E. (2003) The influence of landscape characteristics on duck nesting success in the Missouri Coteau Region of North Dakota. PhD Dissertation. Montana State University.

Sturdivant, R. X., Rotella, J. J., and Russell, R. E. (2007) A smoothed residual based goodness-of-fit statistic for nest-survival models. *Studies in Avian Biology*, **34**, 45-54.

Williams, B. K., Nichols, J. D., and Conroy, M. J. (2002) *Analysis and Management of Animal Populations: modeling, estimation, and decision making*. Academic Press, New York.