

PART FIVE

ECOLOGICAL MISCELLANEA

The first four parts of this book present statistical methods that ought to be part and parcel of every ecologist's repertoire for measuring ecological systems. But there are many other statistical problems in ecology, and we now enter the gray zone of things you might wish to know ecologically and things you should have been taught by your statistics professor.

Population ecologists must estimate both survival and reproduction if they are to understand the demographic equation. Reproductive parameters are species- or group-specific, so the statistical methods for measuring reproductive output in Cladocera are quite different from those for measuring the reproductive rate of deer. For this reason I do not try to discuss the specific methods for estimating reproductive rates in this book.

But survival rates have some general methodological problems in their estimation, and in Chapter 15 I discuss some of the methods for estimating survival that can be applied to many types of animals and plants. There are some idiosyncratic choices even here — for example, I have not found a good discussion of the problem of estimating survival from radio-telemetry data, so I have presented a synopsis here.

Some statistical procedures like transformations are so obvious to statisticians that they are covered all too quickly in most statistics texts. Other newer statistical methods like the bootstrap are most useful to ecologists but are often deemed too sophisticated for introductory statistics courses. In Chapter 16 I present a few of the statistical methods that you may need to know before you take on the real world with all its problems.

CHAPTER 15**ESTIMATION OF SURVIVAL RATES**

(Version 4, 29 January 2014)

Page

15.1 FINITE AND INSTANTANEOUS RATES.....	656
15.2 ESTIMATION FROM LIFE TABLES	660
15.2.1 Methods of Collecting Life Table Data.....	663
15.2.2 Expectation of Further Life.....	670
15.3 ESTIMATION OF SURVIVAL FROM AGE COMPOSITION.....	674
15.4 RADIO-TELEMETRY ESTIMATES OF SURVIVAL.....	678
15.4.1 Maximum Likelihood Method	679
15.4.2 Kaplan-Meier Method	685
15.5 ESTIMATION OF BIRD SURVIVAL RATES.....	688
15.6 TESTING FOR DIFFERENCES IN SURVIVAL RATES.....	690
15.6.1 Log-rank Test.....	690
15.6.2 Likelihood Ratio Test	694
15.6.3 Temporal Differences in Mortality Rates.....	696
15.7 SUMMARY	696
SELECTED READINGS.....	697
QUESTIONS AND PROBLEMS.....	698

Survival is a critical population process, and the estimation of survival rates is one important part of measuring the parameters of the demographic equation. There are many different approaches to estimating survival rates, and we have already discussed the mark-recapture method in Chapter 2 (page 000). In this chapter I will discuss some of the more general techniques for estimating survival rates. There is a large literature on the estimation of survival rates in wild populations and in this chapter we only scratch the surface of the methods available. There is a bias in the methods discussed here toward the larger vertebrates.

15.1 FINITE AND INSTANTANEOUS RATES

The simplest measures of survival are obtained by following a group of individuals over time. This gives us the *finite* survival rate:

$$\begin{aligned} \text{Finite survival rate} &= \frac{\text{Number of individuals alive at end of time period}}{\text{Number of individuals alive at start of time period}} \\ \hat{S}_0 &= \frac{N_t}{N_0} \end{aligned} \quad (15.1)$$

If the time interval is one year, this will give a finite *annual* survival rate. Finite survival rates can range from 0 to 1, and they always apply to some specific time period.

One common problem in calculating survival is to convert observed rates to a standardized time base. For example, you might observe a 0.95 finite survival rate over 38 days and wish to express this as a 30-day rate or as an annual rate. This is easily done with finite rates as follows:

$$\left\{ \begin{array}{l} \text{Adjusted finite} \\ \text{survival rate} \end{array} \right\} = \left(\begin{array}{l} \text{Observed finite} \\ \text{survival rate} \end{array} \right)^{\left(\frac{t_s}{t_0} \right)} \quad (15.2)$$

where t_s = Standardized time interval (e.g. 30 days)
 t_0 = Observed time interval (e.g. 38 days)

Box15.1 illustrates this conversion procedure. In general one should not extrapolate time periods too much when doing these conversions. If you measure survival rates over 3 days it is somewhat misleading to convert this to an annual survival rate.

Box 15.1 CONVERSION OF SURVIVAL RATES TO A STANDARD TIME BASE

1. Keith *et al.* (1984) observed in snowshoe hares a finite survival rate of 0.384 for a time span of 346 days. To convert this survival rate to a standard year (365 days) proceed as follows:

Using equation (15.2):

$$\begin{aligned} \left\{ \begin{array}{l} \text{Adjusted annual} \\ \text{survival rate} \end{array} \right\} &= \left(\begin{array}{l} \text{Observed finite} \\ \text{survival rate} \end{array} \right)^{\left(\frac{t_s}{t_0} \right)} \\ &= (0.384)^{365/346} = (0.384)^{1.0549} = 0.364 \end{aligned}$$

The equivalent annual survival rate for these hares is thus 36.4%.

To express this survival rate as a weekly survival rate:

$$\left\{ \begin{array}{l} \text{Adjusted weekly} \\ \text{survival rate} \end{array} \right\} = \left(\begin{array}{l} \text{Observed finite} \\ \text{survival rate} \end{array} \right)^{\left(\frac{t_s}{t_0} \right)}$$

$$= (0.384)^{7/346} = (0.384)^{0.10693} = 0.903$$

The equivalent weekly survival rate is thus 90.3% in this example.

2. Perennial lupine plants in the southern Yukon suffer a mortality rate of 22% per year. What fraction of these plants will survive for 10 years if the mortality rate does not change? For this example:

$$\left\{ \begin{array}{l} \text{Adjusted 10 year} \\ \text{survival rate} \end{array} \right\} = \left(\begin{array}{l} \text{Observed one year} \\ \text{survival rate} \end{array} \right)^{\left(\frac{t_s}{t_0} \right)}$$

$$= (1.0 - 0.22)^{10/1} = (0.78)^{10} = 0.083$$

or about 8% of plants would live for 10 years in this population.

These calculations can be carried out by Program SURVIVAL (Appendix 2, page 000)

Survival rates can be expressed as finite rates or as instantaneous rates. If the number of deaths in a short time period is proportional to the total number of individuals at that time, the rate of drop in numbers can be described by the geometric equation:

$$\frac{dN}{dt} = iN \quad (15.3)$$

where N = number of individuals
 i = Instantaneous mortality rate (always a negative number)
 t = Time

Figure 15.1 illustrates geometric population decline.

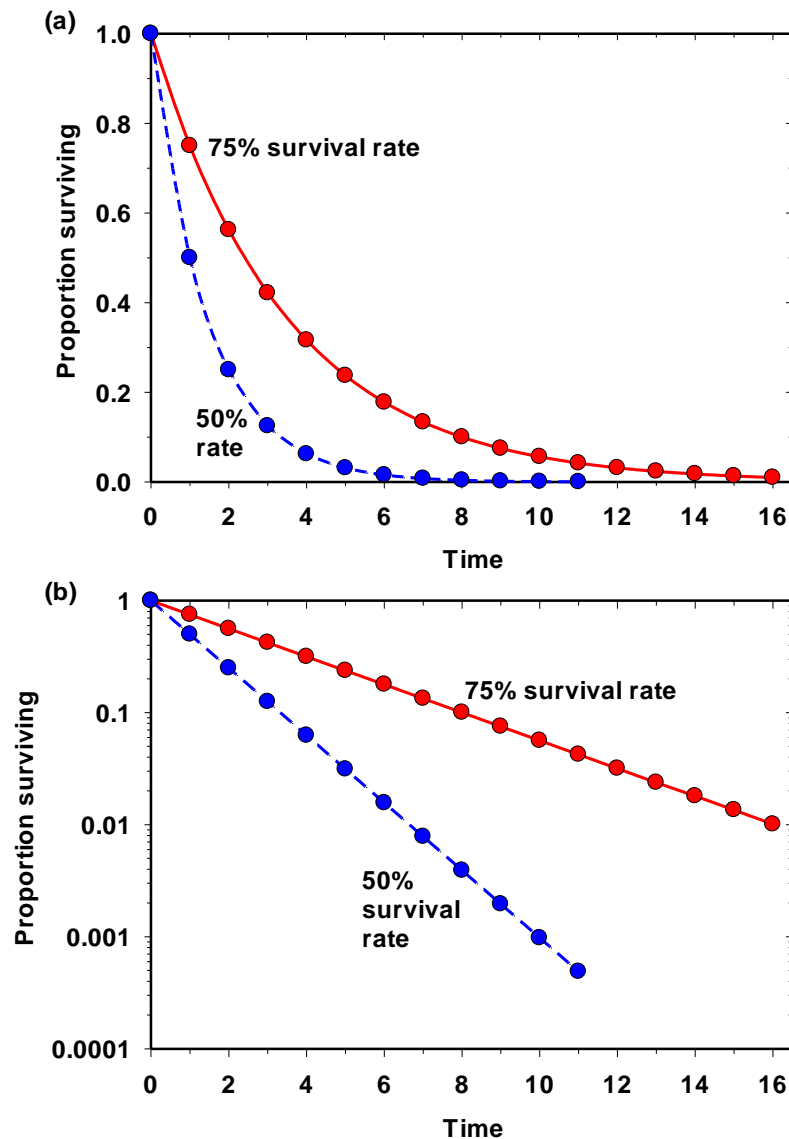


Figure 15.1 Illustration of geometric population decline. Many survival analyses assume that a cohort of animals or plants decreases at a constant survival rate. (a) Geometric declines at 50% per time period and 75%. (b) On a semi-logarithmic plot, in which the proportion surviving is expressed on a log-scale, these same declines are linear. The slope of these lines is the instantaneous mortality rate defined in equation (15.3). This graph illustrates why *geometric* and *logarithmic* are used interchangeably to describe proportional changes in abundance.

In integral form equation (15.3) becomes:

$$N_t = N_0 e^{j t} \quad (15.4)$$

where N_t = Number of individuals at time t

N_0 = Number of individuals at start of time period

If $t = 1$ time unit and we take logs, we obtain:

$$\log_e \left(\frac{N_t}{N_0} \right) = i \quad (15.5)$$

Since (N_t / N_0) is the finite survival rate (from equation 15.1), we have obtained the simple relationship:

$$\log_e (\text{finite survival rate}) = \text{Instantaneous mortality rate} \quad (15.6)$$

Note that there is always some *time unit* that must be specified for both finite and instantaneous rates. Instantaneous mortality rates are always negative numbers because they describe the slope of the lines illustrated in Figure 15.1. Because the finite mortality rate is defined as:

$$\text{Finite mortality rate} = (1.0 - \text{finite survival rate}) \quad (15.7)$$

we can convert between finite and instantaneous rates of mortality by the equation:

$$\text{Finite mortality rate} = 1.0 - e^{\text{instantaneous mortality rate}} \quad (15.8)$$

The following table illustrates the conversion between these three measures of survival:

Finite survival rate	Finite mortality rate	Instantaneous mortality rate
1.0	0.0	0.0
0.95	0.05	-0.051
0.90	0.10	-0.105
0.75	0.25	-0.287
0.50	0.50	-0.693
0.25	0.75	-1.386
0.10	0.90	-2.303
0.01	0.99	-4.605

Note that finite rates of survival and mortality can range from 0 to 1. Instantaneous rates of mortality can range from 0 to $-\infty$. It is important to remember that both finite

and instantaneous rates of survival and mortality always refer to a specific time period such as one month or one year.

It is clearly easiest to think about finite survival rates or finite mortality rates, and survival data should usually be reported in this form. Instantaneous rates are useful mathematically, but they are difficult to conceptualize. Survival is a multiplicative process and one obtains average survival rates by chain multiplication. For example, for an insect population to calculate total generation survival:

$$S_G = S_1 \times S_2 \times S_3 \times S_4 \times S_5 \quad (15.9)$$

where

- S_G = Generation survival rate
- S_1 = Survival of egg stage
- S_2 = Survival of small larval stage
- S_3 = Survival of large larval stage
- S_4 = Survival of pupal stage
- S_5 = Survival of adult stage

The same general principle applied to any type of survival average. The important point is that finite survival rates are not averaged arithmetically but as a geometric mean (as in equation 15.9).

15.2 ESTIMATION FROM LIFE TABLES

The mortality schedule operating on a population can be described succinctly by a *life table*. A life table is an age-specific summary of the mortality rates operating on a population. A population may often be usefully subdivided into groups like males and females, and different life tables may be calculated for each group. Table 15.1 gives an example of a life table for female African buffalo.

Life tables have been presented in several formats and the following symbols are commonly used to summarize the data in the columns of a life table:

x = Age interval

n_x = Number of individuals of a cohort alive at the *start* of age interval x

l_x = Proportion of individuals surviving at the start of age interval x

d_x = Number of individuals of a cohort dying *during* the age interval x to

$$x+1$$

q_x = Finite rate of mortality during the age interval x to $x + 1$

p_x = Finite rate of survival during the age interval x to $x + 1$

e_x = Mean expectation of life for individuals alive at *start* of age x

You should be cautioned that not every ecologist uses these symbols in the same way.

TABLE 15.1 LIFE TABLE FOR FEMALE AFRICAN BUFFALO (*Syncerus caffer*) IN THE SERENGETI REGION OF EAST AFRICA

Age, x (years)	No. alive at start of age interval, n_x	Proportion surviving at start of age interval, l_x	No. dying within age interval x to $x+1$, d_x	Finite rate of mortality, q_x	Finite rate of survival, p_x
0	1000	1.000	330	0.330	0.670
1	670	0.670	94	0.140	0.860
2	576	0.576	11	0.019	0.981
3	565	0.565	12	0.021	0.979
4	553	0.553	21	0.038	0.962
5	532	0.532	16	0.030	0.970
6	516	0.516	29	0.056	0.944
7	487	0.487	20	0.041	0.959
8	467	0.467	35	0.075	0.925
9	432	0.432	52	0.120	0.880
10	380	0.380	44	0.116	0.884
11	336	0.336	73	0.217	0.783
12	263	0.263	67	0.255	0.745
13	196	0.196	56	0.286	0.714
14	140	0.140	49	0.350	0.650
15	91	0.091	43	0.473	0.527
16	48	0.048	26	0.542	0.458
17	22	0.022	15	0.682	0.318
18	7	0.007	7	1.000	0.000
19	0	0.000	-	-	-
Source: Sinclair, 1977.					

To set up a life table you must first decide on the time interval for the data.

With longer-lived animals and plants this is often 1 year, but for voles it might be 1

week and for some insects 1 day. By making the time interval shorter you will increase the detail of the mortality picture shown by the life table. But the shorter the time interval, the more data you will need for the life table.

The first point to be noted about life tables is that, given any one of the columns, you can calculate all the rest. There is nothing "new" in each of the 6 columns of the life table. They are just different ways of viewing the same data. The essential conversion formulae are as follows:

$$n_{x+1} = n_x - d_x \quad (15.10)$$

$$q_x = \frac{d_x}{n_x} \quad (15.11)$$

$$l_x = \frac{n_x}{n_0} \quad (15.12)$$

$$e_x = \frac{T_x}{n_x} \quad (15.13)$$

where

$$T_x = \sum_{i=0}^x L_i$$

$$L_x = \frac{(n_x + n_{x+1})}{2}$$

and all other symbols are defined as above. Table 15.2 combines these equations to illustrate the relationships among the life table variables.

TABLE 15.2 FORMULAS FOR CONVERTING BETWEEN DIFFERENT COLUMNS OF THE LIFE TABLE. TO USE THIS TABLE DETERMINE WHAT TYPE OF DATA YOU HAVE (LEFT SIDE OF TABLE) AND MOVE ACROSS THE ROW TO THE COLUMN OF THE TYPE OF DATA YOU WOULD LIKE TO HAVE.

	n_x	l_x	d_x	q_x	p_x
n_x	n_x	$\frac{n_x}{n_0}$	$n_{x+1} - n_x$	$\frac{n_{x+1}}{n_x} - 1$	$\frac{n_{x+1}}{n_x}$
l_x	$l_x n_0$	l_x	$(l_{x+1} - l_x) n_0$	$1 - \frac{l_{x+1}}{l_x}$	$\frac{l_{x+1}}{l_x}$

d_x	$\sum_{y=x}^{\infty} d_y$	$\frac{\sum_{y=x}^{\infty} d_y}{n_0}$	d_x	$\left(\frac{d_x}{\sum_{y=x}^{\infty} d_y} \right)$	$1 - \left(\frac{d_x}{\sum_{y=x}^{\infty} d_y} \right)$
q_x	$n_0 \prod_{y=0}^{x-1} (1 - q_y)$	$\frac{\prod_{y=0}^{x-1} (1 - q_y)}{n_0}$	$q_x \prod_{y=0}^{x-1} (1 - q_y)$	q_x	$1 - q_x$
p_x	$n_0 \prod_{y=0}^{x-1} p_y$	$\frac{\prod_{y=0}^{x-1} p_y}{n_0}$	$1 - q_x \prod_{y=0}^{x-1} (1 - q_y)$	$1 - p_x$	p_x

Source: Caughley, 1977a.

15.2.1 Methods of Collecting Life Table Data

At least six types of data can be collected to calculate a life table. Caughley (1977a) has reviewed these methods and cautions the ecologist to pick the method of maximal efficiency for the organism being studied.

Method 1: Age at Death Recorded Directly

The number of individuals dying in successive intervals of time is recorded for a group of individuals born at the same time. This is the most precise type of data available because it is based on a single cohort followed through time. Every individual must be known to be alive or dead at each time period. The observed data are the d_x column of the life table and the other columns can be calculated from the formulas in Table 15.2.

Method 2: Cohort Size Recorded Directly

The number of individuals alive in successive intervals of time is recorded for a cohort of individuals. These data are similar to those obtained with Method 1 except that those *surviving* are tallied, not those dying. These data are also precise and specific to the cohort studied. The observed data are the n_x column of the life table, and the other parameters of the life table follow from Table 15.2.

Method 3: Age at Death Recorded for Several Cohorts

Individuals are marked at birth and their age at death recorded, as in Method 1. But several different cohorts are pooled from different years or seasons. These data are usually treated as if the individuals were members of one cohort and the analysis of Method 1 is applied.

Method 4: Age Structure Recorded Directly

The number of individuals aged x in a population is compared with the number of these that die before reaching age $x + 1$. The number of deaths in that age interval, divided by the number alive at the start of the age interval, gives an estimate of q_x directly. Note that the numbers must be counted directly, and the same result cannot be achieved by estimating the proportions in different age classes.

Each of these four methods is completely general and requires no assumptions to be made about population stability or stability of the age structure. The next 2 methods are much more specific because they require you to know the rate of increase of the population and that the age distribution is stable.

Method 5: Ages at Death Recorded For a Population With a Stable Age Distribution and Known Rate of Increase

Often it is possible to find skulls or other remains that give the age at death of an individual. These data can be tallied into a frequency distribution of deaths and thus give d_x directly. To correct for the fact that the population is growing (or declining), each d_x value is corrected as follows:

$$d'_x = d_x e^{rx} \quad (15.14)$$

where

$d'_x =$

Corrected estimate of number dying within the age interval x to $x + 1$

$d_x =$ Observed number of deaths within the age interval x to $x + 1$

$r =$ Instantaneous rate of population growth

$X =$ Age class (0, 1, 2 ...)

$e = 2.71828...$

The life table can then be calculated from Table 15.2 using the corrected d'_x values. If the population is constant, note that e^{rx} will be 1.0 so no correction is needed. Box 15.2 illustrates these calculations.

Box 15.2 LIFE TABLE CONSTRUCTION FOR METHOD 5: AGE AT DEATH RECORDED IN A GROWING POPULATION

Sinclair (1977) picked up skulls of male African buffalo in the southern Serengeti and obtained these data:

Age (years)	No. of skulls found of this age class
2	2
3	5
4	5
5	6
6	18
7	17
8	20
9	17
10	15
11	16
12	18
13	15
14	14
15	8
16	5
17	1
18	0
19	1
<hr/>	
183	

Sinclair (1977) estimated that 48.5% of the male buffalo died in their first year and 12.9% of the original cohort died in their second year, but these young skulls disintegrated and so were not represented in the sample. The simplest way to add these two ages into the observed data is to construct a hypothetical cohort of 1000 skulls:

Age x	Hypothetical cohort of 1000 skulls	Corrected estimate of number dying within the age interval x to $x+1$, d_x'
0	485	485
1	129	139.3
2	4.2	4.9
3	10.5	13.2
4	10.5	14.3

5	12.7	18.7
6	38.0	60.3
7	35.9	61.5
8	42.2	78.1
9	35.9	71.8
10	31.6	68.2
11	33.7	78.6
12	38.0	95.7
13	31.6	86.0
14	29.5	86.7
15	16.9	53.6
16	10.5	36.0
17	2.1	7.8
18	0.0	0.0
19	2.1	9.1
Totals	999.9	1468.8

The original data for ages 2-19 are multiplied by the constant 2.109 to make the new cohort sum to 1000. This hypothetical cohort must now be corrected for the fact that this buffalo population has been increasing at $r = 0.077$ per year. From equation (15.14):

$$d'_x = d_x e^{r \cdot x}$$

$$d'_0 = 485 e^{(0.077(0))} = 485$$

$$d'_1 = 129 e^{(0.077(1))} = 139.3$$

$$d'_2 = 4.2 e^{(0.077(2))} = 4.9$$

and these estimates are listed in the table above. We can now treat these corrected estimates (d'_x) as d_x estimates and calculate the life table in the usual way from the formulas in Table 15.2. For example, to calculate the survivorship curve (l_x) function:

$$l_x = \frac{\sum_{y=x}^{\infty} d_y}{n_0}$$

$$l_0 = \frac{1468.8}{1468.8} = 1.000$$

$$l_1 = \frac{983.8}{1468.8} = 0.670$$

$$l_2 = \frac{844.5}{1468.8} = 0.575$$

and the final results are tabulated below:

Age, x	Proportion surviving to start of age x , l_x
0	1.000
1	0.670
2	0.575
3	0.572
4	0.563
5	0.553
6	0.540
7	0.499
8	0.457
9	0.404
10	0.355
11	0.308
12	0.254
13	0.189
14	0.130
15	0.070
16	0.034
17	0.011
18	0.006
19	0.006

Program SURVIVAL (Appendix 2, page 000) can do these calculations.

Method 6: Age Distribution Recorded For a Population With a Stable Age Distribution and Known Rate of Increase

In this case the age distribution of the population is measured directly by sampling. The number of individuals born is calculated from fecundity rates. The rate of increase of the population must be known. In this case the l_x schedule is estimated directly:

$$l_x = \frac{n_x e^{rx}}{n_0} \quad (15.15)$$

where l_x = Proportion surviving at start of age x
 n_x = Number observed in age class x to $x+1$
 n_0 = Number of births estimated from fecundity data
 r = Instantaneous rate of population growth

Box 15.3 illustrates the use of Method 6.

Box 15.3 LIFE TABLE CONSTRUCTION FOR METHOD 6: AGE DISTRIBUTION RECORDED IN A GROWING POPULATION

Caughley (1970) obtained a sample of Himalayan thar in New Zealand by shooting and got these results:

Age, x (yr)	Sampled age distribution, n_x	Correction factor, e^{rx} ($r =$ 0.12)	Corrected frequency $n_x e^{rx}$	Smoothed frequency, n'_x	Estimates of life table parameters		
					l_x	d_x	q_x
0	43	1.00	43.0	43	1.00	0.37	0.37
1	25	1.13	28.3	27	0.63	0.02	0.03
2	18	1.27	22.9	26	0.61	0.03	0.04
3	18	1.43	25.8	25	0.58	0.03	0.05
4	19	1.62	30.8	23	0.55	0.05	0.09
5	11	1.82	20.0	22	0.50	0.05	0.10
6	12	2.05	24.7	19	0.45	0.06	0.13
7	8	2.32	18.6	17	0.39	0.06	0.15
8	2	2.61	5.2	14	0.33	0.06	0.18
9	3	2.94	8.8	11	0.27	0.06	0.21
10	4	3.32	13.3	9	0.21	0.05	0.23
163							

The age distribution at birth (age 0) was estimated from fecundity rates. The analysis proceeds as follows:

1. Calculate the corrected age distribution from the equation:

$$n'_x = n_x e^{rx}$$

where n'_x = Corrected number of individuals in age class x
 n_x = Observed number of individuals in age class x
 r = Instantaneous rate of population growth

For example:

$$n'_3 = 18 e^{(0.12)(3)} = 25.8$$

$$n'_6 = 12 e^{(0.12)(6)} = 24.7$$

These values are given in column 4 of the table above.

2. Smooth the corrected age distribution either by some smoothing function (moving averages, splines) or by eye. This step is necessary because the age distribution must *decline* with age (i.e. a monotonic, decreasing function). For example, it is logically impossible for there to be more 3-year-olds than 2-year-olds if the population is growing smoothly. In this particular case, smoothing was done by eye, and the results are given in column 5 of the table above.

3. The smoothed age distribution can now be used as n_x values to calculate the life table functions, as in Table 15.2. For example, to calculate the survivorship function (l_x) from equation (15.12):

$$l_x = \frac{n_x}{n_0}$$

Thus we obtain:

$$l_1 = \frac{27}{43}$$

$$l_2 = \frac{26}{43} = 0.605$$

$$l_3 = \frac{25}{43} = 0.581$$

and these values are given in the above table.

Program SURVIVAL (Appendix 2, page 000) can do these calculations from a smoothed age distribution.

Methods 5 and 6 both assume that the rate of population growth is known and that the age distribution is stable. Caughley (1977a) cataloged the numerous ways in which life tables have been calculated incorrectly in the literature by means of Methods 5 and 6. Most of these pitfalls involve biased sampling. For example, if only winter-kills are located, there are no data on summer mortality and a life-table cannot be computed. If the conspicuousness of carcasses varies with age, the observed ages at death are a biased frequency distribution and should not be converted to a life table. Animals shot by hunters often represent a very biased age distribution.

A life table calculated from age distribution data like Method 6 will give appropriate estimates only when the frequency of each age class is the same as or less than that of the preceding age class. If this is not the case, the d_x values will be negative, which is impossible. In these cases Caughley (1977a) recommends smoothing the age distribution. Many computer packages for data analysis and statistics have smoothing functions that can also be used to smooth age distributions. Cleveland (1994) illustrates the use of cubic splines to smooth a data series.

There is little discussion in the literature about the sample sizes needed for the construction of a life table. Caughley (1977a) recommends at least 150 individuals as a minimum when age distributions are used in the analysis. Age determinations may also be crude in many species and if age is not measured carefully, the resulting life table may be very inaccurate. I have seen no discussion of power analysis applied to life table data.

15.2.2 *Expectation of Further Life*

Life table data may be summarized in yet another form by calculating the expectation of further life. The mean expectation of further life can be used as one way of compressing an entire life table into one number. But this compression is done at a price of leaving out the details of the life table that may be of greatest interest to an ecologist.

To calculate mean expectation of life, proceed as follows:

1. Calculate the *life-table age structure* L_x :

$$L_x = \frac{n_x + n_{x+1}}{2} \quad (15.16)$$

where L_x = Number of individuals alive on average during the age interval
 x to $x+1$
 n_x = Number of individuals alive at start of age interval x

This value is only approximate if numbers do not fall linearly within the age interval, but for most animal and plant populations this is a reasonable approximation. If the

time interval is long or mortality rates are high, a better estimate of L_x might be the geometric mean:

$$L_x = \sqrt{n_x(n_{x+1})} \quad (15.17)$$

For example, if $n_3 = 50$ and $n_4 = 10$, equation (15.16) would give $L_3 = 30$ while equation (15.17) would give $L_3 = 22.4$.

2. Calculate the for each age class T_x :

$$T_x = \sum_{i=x}^m L_i \quad (15.18)$$

where T_x = A cumulative function of (individuals)(time units) for age x
 individuals
 L_x = Life-table age structure at age x
 m = Maximum age class observed

3. Estimate the *mean expectation of further life* from the equation:

$$e_x = \frac{T_x}{n_x} \quad (15.19)$$

where e_x = Mean expectation of further life for individuals alive at start of
 age interval x
 T_x = Cumulative function of (individuals)(time units) for age x
 (equation 15.18)
 n_x = Number of individuals alive at start of age interval x

Leslie *et al.* (1955) point out that the special case of the mean expectation of life for an individual alive at the start of the life table is approximately:

$$\hat{e}_0 = 0.5 + l_1 + l_2 + l_3 + \dots + l_{m-1} + 0.5l_m \quad (15.20)$$

where \hat{e}_0 = Mean expectation of life for an individual alive at the start of
 the life table
 l_x = Proportion of individuals surviving to the start of age interval x
 l_m = Proportion of individuals surviving to the start of the last age
 interval m

The variance of the estimate e_0 is given by Leslie *et al.* (1955) as:

$$\text{var}(\hat{e}_0) = \sum_{x=0}^{m-1} \left[\frac{S_{x+1}^2 q_x}{p_x(n_x - 0.5a_x)} \right] \quad (15.21)$$

where $\text{var}(\hat{e}_0)$ = Variance of the mean expectation of life for an individual
 alive at the origin of the life table
 q_x = Finite rate of mortality during the age interval x to $x+1$
 (equation 15.10)
 p_x = Finite rate of survival during the age interval x to $x+1 = 1.0 - q_x$
 n_x = Number of individuals known to be alive at start of age interval
 x
 a_x = Number of accidental deaths (or removals) during the interval x
 to $x+1$
 $S_x = l_x + l_{x+1} + \dots + l_{m-1} + 0.5 l_m$
 l_x = Proportion of individuals surviving to the start of age interval x
 m = Number of age groups in the data

From this variance you can obtain confidence intervals in the usual way:

$$\hat{e}_0 \pm t_\alpha (S.E.) \quad (15.22)$$

where \hat{e}_0 = Mean expectation of life from age 0
 t_α = Student's t -value for $(n-1)$ degrees of freedom
 $S.E.$ = Standard error of the estimate $\hat{e}_0 = \sqrt{\text{var}(\hat{e}_0)}$ given above

Box 15.4 illustrates these calculations for expectation of life.

Box 15.4 CALCULATION OF EXPECTATION OF FURTHER LIFE IN SNOWSHOE HARES

A cohort of 63 snowshoe hares was followed at 6-month intervals with the following results:

Age class, x	Age (months)	No. alive, n_x	Proportion alive, l_x	Mortality rate, q_x	Accidental deaths, a_x
0	0	63	1.00	0.33	0
1	6	42	0.67	0.14	0
2	12	36	0.57	0.33	1
3	18	24	0.38	0.38	0
4	24	15	0.24	0.53	0
5	30	7	0.11	0.71	0
6	36	2	0.03	1.00	0
7	42	0	0.00	-	0

The first step is to calculate the L_x , T_x , and S_x functions needed to solve equations (15.16) to (15.19). Fill in this table as follows:

Age class	L_x	T_x	S_x
-----------	-------	-------	-------

0	52.5	157.5	2.9850
1	39	105	1.9850
2	30	66	1.3183
3	19.5	36	0.7469
4	11	16.5	0.3659
5	4.5	5.5	0.1279
6	1	1	0.0167

From equation (15.16):

$$L_x = \frac{n_x + n_{x+1}}{2}$$

$$\hat{L}_0 = \frac{63 + 42}{2} = 52.5$$

$$\hat{L}_1 = \frac{42 + 36}{2} = 39$$

and so on. These results are listed above.

From equation (15.18)

$$T_x = \sum_{x=0}^{\infty} L_x$$

$$\hat{T}_0 = 52.5 + 39 + 30 + 19.5 + 11 + 4.5 + 1 = 157.5$$

$$\hat{T}_1 = 39 + 30 + 19.5 + 11 + 4.5 + 1 = 105$$

and so on. These results are listed in the table above.

$$S_x = l_x + l_{x+1} + l_{x+2} + \cdots + l_{m-1} + 0.5l_m$$

$$\hat{S}_0 = 1.00 + 0.67 + 0.57 + 0.38 + 0.24 + 0.11 + (0.5)(0.03) = 2.985$$

and so on, with the results listed in the table above..

From these summations the mean expectation of life at birth follows from equation (15.19):

$$e_0 = \frac{T_0}{n_0} = \frac{157.5}{63} = 2.5 \text{ time units}$$

Since one time unit in this example is 6 months, this can be converted directly to:

$$\hat{e}_0 = (2.5)(6) = 15 \text{ months}$$

The variance of this estimate from equation (15.21) is:

$$\begin{aligned}
 \text{var}(\hat{e}_0) &= \sum_{x=0}^{m-1} \left[\frac{S_{x+1}^2 q_x}{p_x (n_x - 0.5a_x)} \right] \\
 &= \frac{(1.985^2)(0.3333)}{0.6667(63 - 0)} \\
 &\quad + \frac{(1.3183^2)(0.1429)}{0.0857(42 - 0)} \\
 &\quad + \frac{(0.7469^2)(0.3333)}{0.6667(36 - 0.5)} + \dots = 0.0503
 \end{aligned}$$

The standard error of the estimate of e_0 is:

$$\begin{aligned}
 S.E. &= \sqrt{\text{var}(e_0)} \\
 &= \sqrt{0.0503} = 0.2244 \text{ time units}
 \end{aligned}$$

or expressed in months:

$$S.E. = (0.244)(6) = 1.346 \text{ months}$$

The 95% confidence limits are obtained from Student's t -distribution with 62 degrees of freedom:

$$\begin{aligned}
 &\hat{e}_0 \pm t_{\alpha} (S.E.) \\
 &15.0 \pm 2(1.346)
 \end{aligned}$$

or 12.3 to 17.7 months for these hares.

Program SURVIVAL (Appendix 2, page 000) can do these calculations.

15.3 ESTIMATION OF SURVIVAL FROM AGE COMPOSITION

Fisheries scientists have developed a whole series of methods for estimating survival rates from data on age composition. Ricker (1975, Chapter 2) gives a detailed exposition of these methods, which I will summarize here.

When it is possible to age individual organisms, like fish or trees, one can estimate survival rates directly from the ratio of numbers in each successive age group:

$$\hat{S}_t = \frac{N_{t+1}}{N_t} \quad (15.23)$$

where S_t = Finite annual survival rate of individuals in age class t
 N_{t+1} = Number of individuals in age class $t+1$
 N_t = Number of individuals in age class t

This simple approach to estimating survival can be used only when three assumptions are satisfied:

- The survival rate is constant for each age group
- All year-classes are recruited at the same abundance
- All ages are sampled equally by the sampling gear

These assumptions are very restrictive and cannot be correct for many populations.

If the survival rate is constant over a period of time, a combined estimate of average survival can be made from formulas given by Robson and Chapman (1961):

$$\hat{S} = \frac{T}{R + T - 1} \quad (15.24)$$

where \hat{S} = Finite annual survival rate estimate
 $T = N_1 + 2N_2 + 3N_3 + 4N_4 + \dots$
 $R = \sum_{t=0}^m N_t$
 N_t = Number of individuals in age group t

This estimate of survival has a variance of:

$$\text{var}(\hat{S}) = \hat{S} \left[\hat{S} - \left(\frac{T-1}{R+T-2} \right) \right] \quad (15.25)$$

Ricker (1975) gives the following age data on Antarctic fin whales from 1947 to 1953 to illustrate these calculations:

Age	0	1	2	3	4	5	6+
Frequency (%)	0.3	2.3	12.7	17.2	24.1	15.1	29.5

For these whales, ages 0-3 are not sampled representatively by the whaling ships, so these data must be omitted from analysis. From equation (15.27) for ages 4 to 5:

$$\hat{S} = \frac{14.1}{24.1} = 0.585 \text{ per year}$$

A combined estimate from Robson and Chapman's method gives, from equation (15.24):

$$\begin{aligned}\hat{T} &= 14.1 + 2(29.5) = 73.1 \\ \hat{R} &= 24.1 + 14.1 + 29.5 = 67.7 \\ \hat{S} &= \frac{73.1}{67.7 + 73.1 - 1} = 0.523 \text{ per year}\end{aligned}$$

with variance:

$$\text{var}(\hat{S}) = 0.523 \left[0.523 - \left(\frac{73.1 - 1}{67.7 + 73.1 - 2} \right) \right] = 0.00133$$

Fisheries scientists discovered in 1908 that a plot of size-frequency data often formed a dome shaped curve with a long descending right limb. Baranov (1918) called these *catch curves*, and most fishery scientists now apply this name to plots of log frequency-at-age (*y*-axis) against age (*x*-axis). The analysis of catch curves can be useful in fisheries management. Figure 15.2 illustrates a catch curve for the petrale sole off the coast of western Canada.

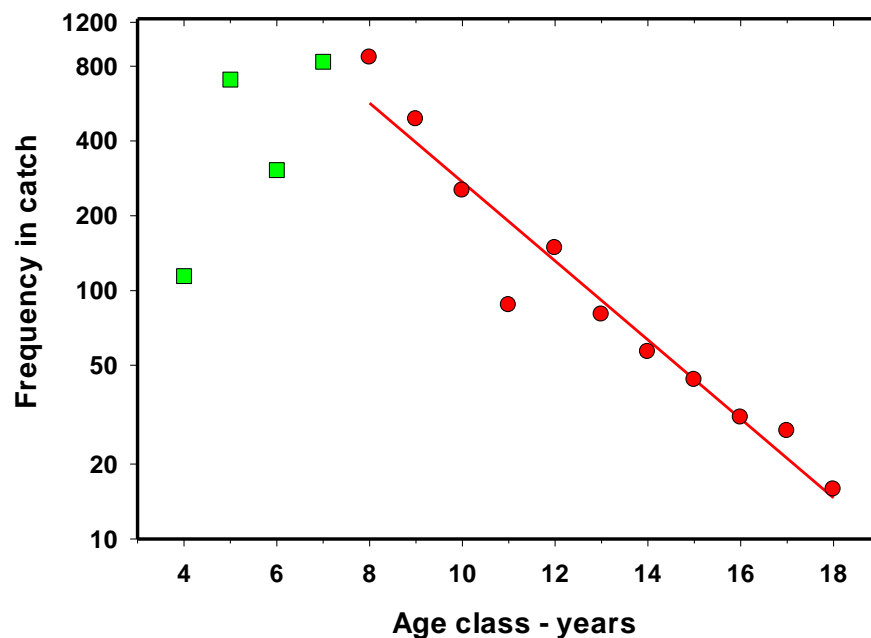


Figure 15.2 Catch curve for the 1945 year class of female petrale sole (*Eopsetta jordani*) off western Canada. The slope of the right limb of the catch curve (ages 8 to 18 years, •) is -0.36 . (Data from Ketchen and Forrester, 1966)

The ascending left part of a catch curve and the dome of the curve represent age groups that are not adequately sampled by the gear. In a commercial fishery young fish may live in different habitats or not be caught in the nets used to fish for the adults. The more interesting part of the curve is the descending right limb.

Baranov (1918) showed that the right limb is a survivorship curve that is both age-specific and time-specific if the following assumptions are correct:

1. The mortality rate is uniform with age
2. The mortality rate is constant over time
3. The sample is taken randomly from the age-groups involved
4. Recruitment is constant for all age groups.

If these assumptions are correct, the finite rate of survival can be estimated from the antilog of the slope of the right limb of the catch curve (for the linear regression $Y = \log_e (\text{frequency-at-age})$ and $x = \text{age class}$). Alternatively, for two age classes t and $t+1$, the survival rate is: (using base e logs)

$$i = \log_e (f_t) - \log_e (f_{t+1}) \quad (15.26)$$

$$\hat{S} = e^{-i} \quad (15.27)$$

where f_t = Frequency of age t individuals
 f_{t+1} = Frequency of age $t+1$ individuals

\hat{S} = Estimate of finite survival rate between age t and $t+1$

For example, in Figure 15.2 the slope of the right limb of the curve for \log_e (frequency) versus age is -0.36056. The finite survival rate is thus:

$$\hat{S} = e^{-0.36056} = 0.70 \text{ per year}$$

Some of these assumptions can be relaxed in certain situations. For example, if a fishery is sampled every year, a single year-class can be followed throughout its life span (e.g. the 1975 year class can be sampled in 1984 when it is age IX, in 1985 when it is age X, etc.). In this case the assumption of constant recruitment can be dropped, and the catch curve for that year class is a cohort survivorship curve for which the average survival rate can be calculated.

If these four assumptions are violated, the right limb of the catch curve will often not be a straight line. Ricker (1975) discusses in detail the interpretation of nonlinear catch curves from commercial fisheries.

Variations in the rate of recruitment will usually not make a catch curve nonlinear, but will change the slope of the right limb. For example, if recruitment is steadily falling over time, the resulting catch curve will be linear but have a lower slope, so the estimated survival rate will be too high.

Variations in mortality rates will tend to make the catch curve bend. If mortality rate increases with age, the catch curve will show a convex right limb. If mortality rate decreases with age, a concave right limb will be produced. If the mortality rate varies greatly from year to year, catch curves will not be linear and are not very useful for survival estimation.

If age determination is not accurate, the catch curve will not be affected, as long as there is no bias in the age estimation. In practice there is often a negative bias in age determination — older individuals are usually assigned to younger ages (Ricker 1975). If age determination is strongly biased, estimated survival rates from catch curves are too low. The problem can be overcome by making sure that age determinations are accurate.

15.4 RADIO-TELEMETRY ESTIMATES OF SURVIVAL

One use of radio-collars for the study of animal populations is to estimate survival rates. In studies of this type, radio transmitters are placed on several individuals and they are then followed until death or radio-malfunction or accidental loss of the radio collar. Trent and Rongstad (1974) were the first to discuss methods of estimating survival rates from telemetry data. They proposed the following approach. The average daily survival rate is estimated by:

$$\hat{S} = \frac{x - y}{x} \quad (15.28)$$

where \hat{S} = Estimate of finite daily survival rate

x = Total number of radio-days observed over the period

y = Total number of deaths observed over the period

For example, Trent and Rongstad (1974) observed 31 cottontail rabbits for a total of 1660 radio-rabbit-days in September and October 1971, and in these two months observed 6 deaths. Thus:

$$\hat{S} = \frac{1660 - 6}{1660} = 0.99638 \quad (\text{survival rate per day})$$

To convert these daily rates to weekly, monthly, or annual rates of survival, use the equation:

$$\hat{p} = \hat{S}^n \quad (15.29)$$

where p = Estimate of finite survival rate per n days
 S = Finite daily survival rate (from equation 15.32)
 n = Number of days you wish to convert to

For example, for these cottontails a 28-day survival rate would be

$$\hat{p} = 0.99638^{28} = 0.9036 \quad (\text{per 28 days})$$

This method of estimating survival rates assumes that each rabbit-day is a binomial event in which the animal lives or dies, so that the distribution of deaths should follow a binomial distribution. In order to analyze the data this way, you must assume that each radio-day is an independent event and that there is a constant mortality rate over the whole time period. The method assumes that every death is recorded exactly on the day it occurs without guesswork.

To obtain confidence limits for these estimates, Trent and Rongstad (1974) use the Poisson approximation to the binomial, and thus used the methods outlined in Chapter 4 (page 000). Another approach would be to use the binomial distribution directly to obtain confidence limits (Program EXTRAS, Appendix 2, page 000).

15.4.1 Maximum Likelihood Method

Problems arise with the Trent and Rongstad (1974) method when the exact time of death is not known. Mayfield (1961, 1975) discussed the problem of estimating nest success in birds in which nests are visited periodically. This is exactly the same problem of estimation that Trent and Rongstad faced with radio-collared rabbits. Mayfield (1975) assumed that each nest loss occurred exactly half-way between visits. For example, if a nest was present on day 15 and then found destroyed at the

next visit on day 19, one would assume it failed on day 17. Johnson (1979) recognized that the Mayfield method led to biased estimates of nest success if nests are visited at irregular intervals, and estimated nest success values would be too high. Hensler (1985) further discusses the possible biases of the Mayfield method.

These problems of estimation were resolved by Bart and Robson (1982) who suggested a maximum-likelihood estimator of the survival rate calculated as follows:

1. Compile the survival data in a frequency table like Table 15.3 in which interval length between observations is given for each observation of survival or mortality.

TABLE 15.3 SURVIVAL DATA FOR NESTLING MOURNING DOVES FROM PERIPHERAL NESTING LOCATIONS, TALLIED IN A FREQUENCY DISTRIBUTION FOR CALCULATION OF THE MAXIMUM LIKELIHOOD ESTIMATOR OF FINITE DAILY SURVIVAL RATE^a

Interval between subsequent visits to nests (days) (L)	Total no. of nests sampled with this interval (n_L)	No. of nests surviving intact (n_{LS})	No. of nests with mortality (n_{LF})
1	12	12	0
2	8	7	1
3	15	12	3
4	18	17	1
5	26	24	2
6	33	31	2

^a Data from Bart and Robson, 1982.

2. Calculate the Mayfield (1975) estimator of the daily survival rate:

$$\hat{S} = 1 - \left(\frac{\text{Number of deaths}}{\sum_L L(n_{LS} + 0.5n_{LF})} \right) \quad (15.30)$$

where \hat{S} = Mayfield estimate of finite daily rate of survival

L = Interval length in days (1, 2, 3,...)

n_{LS} = Number of intervals of length L in which no mortality occurred

n_{LF} = Number of intervals of length L in which some mortality

occurred

3. Use the Mayfield estimate \hat{S} as a preliminary estimate for the maximum likelihood estimate from the following equations:

(a) Calculate A and B from:

$$A = \sum_L \left[\frac{L}{\hat{S}} \left(n_{LS} - \frac{n_{Lf} \hat{S}^L}{1 - \hat{S}^L} \right) \right] \quad (15.31)$$

$$B = \sum_L \frac{L}{\hat{S}^2} \left[n_{LS} + \frac{n_{Lf} \hat{S}^L (L - 1 + \hat{S}^L)}{(1 - \hat{S}^L)^2} \right] \quad (15.32)$$

where all terms are defined above.

(b) The maximum likelihood estimate of \hat{S} is given by:

$$\hat{S}_M = \hat{S} + \frac{A}{B} \quad (15.33)$$

where

\hat{S}_M = Maximum likelihood estimate of survival rate

\hat{S} = Mayfield estimate of survival rate

A, B are defined above

Use this estimate as a trial value of \hat{S} , go back to **3** and repeat steps (a) and (b) a second time. Bart and Robson (1982) found that 2 or 3 iterations were adequate for convergence to a final estimate of the daily survival rate. Box 15.5 gives an example of these calculations.

Box 15.5 ESTIMATION OF SURVIVAL RATE FROM RADIO TELEMETRY DATA

Red squirrels were fitted with radio collars and monitored over a 2-month period with these results:

Intervals between relocations (days), L	No. of red squirrels relocated, n_L	No. of survivors, n_{LS}	No. of deaths, n_{Lf}
1	47	45	2
2	23	22	1
3	36	33	3
4	12	12	0

Trent and Rongstad estimator

From equation (15.28):

$$X = \text{Total number of radio-days observed} \\ = (47)(1) + (23)(2) + (36)(3) + (12)(4) = 249 \text{ radio-days}$$

$$\hat{S} = \frac{X - Y}{X} \\ = \frac{249 - 6}{249} = 0.9759 \quad (\text{finite survival rate per day})$$

Mayfield estimator

From equation (15.30):

$$\sum_L L(n_{LS} + 0.5n_{Lf}) = 1[45 + (0.5)(2)] + 2[22 + (0.5)(1)] + \\ 3[33 + (0.5)(3)] + 4[12 + 0] = 242.5 \text{ radio-days}$$

$$\hat{S} = 1 - \left(\frac{\text{number of deaths}}{\sum L(n_{LS} + 0.5n_{Lf})} \right) \\ = 1 - \left(\frac{6}{242.5} \right) = 0.9526 \quad (\text{finite survival rate per day})$$

Bart and Robson (1982) Maximum-likelihood estimator

From equations (15.31), (15.32) and (15.33):

$$A = \sum_L \left[\frac{L}{\hat{S}} \left(n_{LS} - \frac{n_{Lf} \hat{S}^L}{1 - \hat{S}^L} \right) \right] \\ = \frac{1}{0.97526} \left[45 - \frac{2(0.97526)}{1 - 0.97526} \right] + \\ \frac{2}{0.97526} \left[22 - \frac{1(0.97526^2)}{1 - 0.97526^2} \right] + \dots = 3.0183154$$

$$\begin{aligned}
 B &= \sum_L \frac{L}{\hat{S}^2} \left[n_{LS} + \frac{n_{Lf} \hat{S}^L (L-1 + \hat{S}^L)}{(1 - \hat{S}^L)^2} \right] \\
 &= \frac{1}{(0.97526)^2} \left[45 + \frac{2(0.97526)(1-1+0.97526)}{(1-0.97526)^2} \right] + \\
 &\quad \frac{2}{(0.97526)^2} \left[22 + \frac{1(0.97526^2)(1-1+0.97526^2)}{(1-0.97526^2)^2} \right] + \dots = 10,059.89
 \end{aligned}$$

$$\begin{aligned}
 \hat{S}_M &= \hat{S} + \frac{A}{B} \\
 &= 0.97526 + \frac{3.0183154}{10059.89} = 0.975558 \quad (\text{finite daily survival rate})
 \end{aligned}$$

A second iteration using this estimate of 0.975558 for \hat{S} gave:

$$\hat{S}_M = 0.975555$$

and a third iteration gave the same result:

$$\hat{S}_M = 0.975555$$

so there was no need to continue the iterations. This is the best estimate of the daily survival rate for these red squirrels.

These calculations, along with estimates of the 95% confidence intervals, can be carried out by Program SURVIVAL (Appendix 2, page 000)

To obtain confidence limits for this estimate of survival, Bart and Robson (1982) showed that a square root transformation was required. The calculations are somewhat tedious and proceed as follows:

(1) Calculate the transformed survival rate

$$\hat{S}_t = \sqrt{1 - \hat{S}_M^x} \quad (15.34)$$

where

\hat{S}_t = Transformed daily survival rate

\hat{S}_M = Maximum likelihood estimate of survival (eq. 15.35)

x = Average interval length = $\frac{\sum L n_L}{\sum n_L}$

$n_L = n_{LS} + n_{Lf}$ = Total number of observations in interval L

L = Interval length in days

(2) Calculate the standard error of \hat{S}_t :

$$\hat{S}_{pt} = \frac{x(\hat{S}_M^{x-1})\hat{S}_p}{2\sqrt{1-\hat{S}_M^x}} \quad (15.35)$$

where \hat{S}_{pt} = Standard error of the transformed daily survival rate
 x = Average interval length (defined above)
 \hat{S}_M = Maximum likelihood estimate of survival

$$\hat{S}_p = \left[\sum_L \left(\frac{n_L L^2 (\hat{S}_M^{L-2})}{1 - \hat{S}_M^L} \right) \right] \quad (15.36)$$

Program SURVIVAL (Appendix 2, page 000) can do these calculations.

This transformation can also be used to estimate the sample size needed to get a specified precision in an estimate of the daily survival rate. Bart and Robson (1982) give this procedure:

1. Guess the likely value of the finite daily survival rate (S) (e.g. 0.98)
2. Specify the desired lower confidence limit (S_L) (e.g. 0.95) on the assumption that the finite daily survival rate given in (1) is correct
3. Guess the average interval length for your data (A) (e.g. 2 days)
4. Decide on the probability you wish to tolerate that the lower estimate will be *less than* the specified limit (e.g. 0.025)

From these values the sample size you need is given by:

$$\hat{n} = \left(\frac{z S^L}{2 \Delta} \right)^2 \quad (15.37)$$

where n = Estimated total sample size needed
 z = Standard normal deviate for the probability level chosen in (4)
 (e.g. if $p = .025$, z will be 1.96, if $p = .05$, z will be 1.645)
 S = Guessed value of the daily survival rate
 L = Guessed value of the average interval length / 2.0
 $\Delta = \sqrt{1 - S_L^A} - \sqrt{1 - S^A}$
 A = Average interval length (guessed)
 S_L = Desired lower confidence limit for the probability level chosen

To be conservative, you should underestimate the daily survival rate slightly.

One restriction on these estimates is that survival rates are assumed to be constant within the time unit specified. Heisey and Fuller (1985) describe a computer program that allows one to combine data for several time periods and test the hypothesis that survival was constant throughout the whole time period. White (1983) describes a completely general survival program that computes maximum likelihood estimates for any type of radio telemetry, band-recovery, or nest success data. It is important to recognize that survival rates in natural populations are rarely constant from year to year or equal in the two sexes. White (1983) warns against uncritical acceptance of the null hypothesis of equal survival rates when sample sizes are small and the ability to specify narrow confidence limits does not exist. There is no substitute for reasonable sample sizes in any estimation of survival rates.

15.4.2 Kaplan-Meier Method

An alternative approach to estimating survival rates from radio-collared animals has been proposed by Pollock *et al.* (1989). This approach centers on the time at death of each individual in the sample. Individuals may live through the time period of interest or lose their radio during the period. The survival rate from the start of the period until the day of death of the last recorded death is given by:

$$\hat{S}_K = \prod_{i=1}^n \left[1 - \left(\frac{d_i}{r_i} \right) \right] \quad (15.38)$$

where \hat{S}_K = Kaplan-Meier estimate of finite survival rate for the period

d_i = Number of deaths recorded at time i

r_i = Number of individuals alive and at risk at time i

n = Number of time checks for possible deaths

Radio-tagged individuals may be checked every day or week, every second day or week, or at irregular time periods for this method. Table 15.4 illustrates the type of data obtained on radio-tagged bobwhite quail.

TABLE 15.4 SURVIVAL DATA ON RADIO-TAGGED BOBWHITE QUAIL, SUMMARIZED FOR KAPLAN-MEIER SURVIVAL ESTIMATION^a

Sample week(i)	Time of checks	No. birds with radios (r_i)	No. found dead (d_i)	No. censored
--------------------	----------------	---------------------------------	--------------------------	--------------

1	17-23 Nov.	20	0	0
2	24-30 Nov.	21	0	0
3	1-7 Dec.	22	2	1
4	8-14 Dec.	19	5	0
5	15-21 Dec.	14	3	0
6	22-28 Dec.	11	0	0
7	29 Dec.-4 Jan.	11	0	0
8	5-11 Jan.	11	2	0
9	12-18 Jan.	9	1	0
10	19-25 Jan.	8	0	1
11	26 Jan.-1 Feb.	7	0	0
12	2-8 Feb.	10	0	0
13	9-15 Feb.	16	4	0
14	16-22 Feb.	22	4	0
15	23 Feb.-1 March	23	4	1
16	2-8 March	24	4	0
17	9-15 March	20	2	0
^a Source: Pollock <i>et al.</i> (1989), Table 2.				

The Kaplan-Meier method has two important advantages for field ecologists. First, newly-tagged animals may be added to the sample at any time, so that the cohort being studied can be kept to a large size even if deaths are occurring throughout the study. This is called the *staggered entry design* and Table 15.4 illustrates this with data on bobwhite quail. The previous methods all assume a cohort of individuals is marked at one time and the subsequent history of that cohort is studied without any possible additions. It is important in any survival study to keep the sample size as large as possible, so that confidence limits are relatively narrow. Second, animals may be lost to the study inadvertently without affecting the estimated survival rate. These individuals represent *censored data* in statistical jargon, individuals who have dropped out of the study without us knowing their fate (Figure 15.3). Animals may lose their radio tags, or radios may fail electronically. The Kaplan-Meier method accounts for censored data by adjusting the number of

individuals at risk. Note that we do *not* assume that censored animals have died but rather that they go on living without us knowing their subsequent fate. We make use of their data up to the point at which they were lost from the sample.

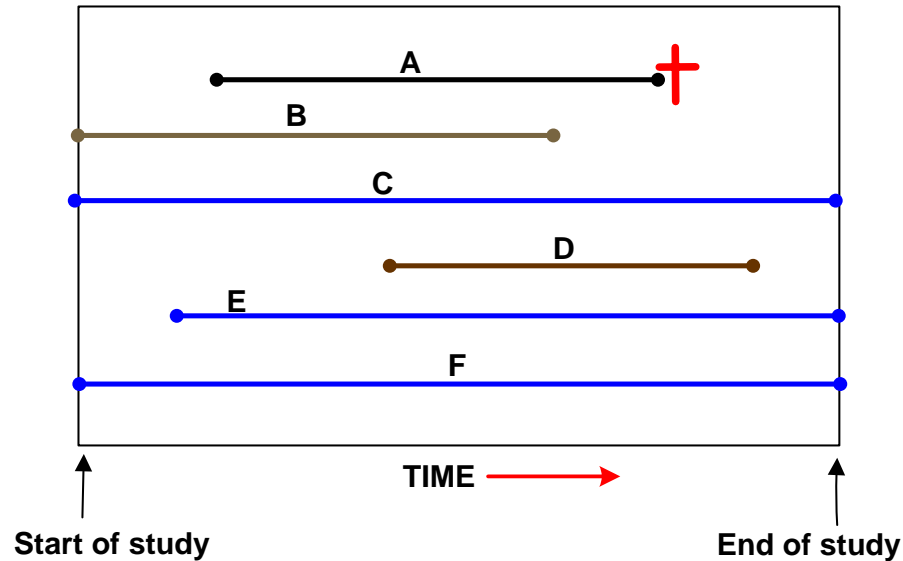


Figure 15.3 Illustration of censored survival data in a mark-recapture or radio-telemetry study. The X-axis represents time which is a defined study period. In this example 6 individuals (A to F) enter the study at different times when they are first captured (the *staggered-entry* design). Individuals alive at the end of the study (C, E, and F in blue) are all *right-censored* since we know they lived that long but not how much longer. During the study animals may lose their tags or radio-collars and thus be lost to observation. In this example animals B and D (brown) are *censored* observations and we use their data to the point of disappearance. One animal (A, black line) died during the study period.

The variance of the Kaplan-Meier estimate of the survival rate can be calculated in two ways and I present here Greenwood's formula (Pollock *et al.* 1989):

$$\text{var}(\hat{S}_K) = \hat{S}_K^2 \left[\sum_{i=1}^n \left(\frac{d_i}{r_i(r_i - d_i)} \right) \right] \quad (15.39)$$

where $\text{var}(\hat{S}_K)$ = Greenwood's estimate of the variance of the Kaplan-Meier survival rate

d_i = Number of deaths recorded at time i

r_i = Number of radio-tagged animals at risk at time i

The standard error of this estimate is the square root of this variance, and confidence limits may be obtained in the usual way:

$$\hat{S}_K \pm t_\alpha \sqrt{\text{var}(\hat{S}_K)} \quad (15.40)$$

The t -value can be based conservatively on the number of radio-tagged individuals alive at the end of the time period or more liberally on the average number of tagged individuals during the study period.

Pollock *et al.* (1989) discuss the advantages of this estimator over the Mayfield (1975) and Bart and Robson (1982) estimators.

Program SURVIVAL (Appendix 12, page 000) can do these calculations.

15.5 ESTIMATION OF BIRD SURVIVAL RATES

A common technique for studying survival in bird populations has been to band young birds in the nest year after year and to record the recoveries of these birds in subsequent years. Table 15.5 illustrates the type of data often obtained in these studies. Anderson *et al.* (1985) and Lebreton *et al.* (1993) have discussed the statistical problems associated with survival data of this type. If the recovery data are from dead birds, there is clearly an incomplete registration of deaths since not all dead tagged birds are located. Similarly, if the recovery data is of survivors, there is an incomplete registration of surviving animals. The probability of survival may vary both in time and with respect to the age of the birds. To separate time- and age-dependence of survival, bird ringing schemes should mark both adult birds and young birds each year (Anderson *et al.* 1985).

TABLE 15.5 RINGING AND RECOVERY DATA FOR YOUNG BIRDS RINGED OVER k YEARS AND RECOVERED OVER l YEARS ($l > k$)^a. THIS APPROACH IS NOT RECOMMENDED FOR SURVIVAL ESTIMATION.

Ringing year	Number ringed	Recoveries (R_{ij}) in year j						
		1	2	3	...	k	...	l
1	N_1	R_{11}	R_{12}	R_{13}	...	R_{1k}	...	R_{1l}
2	N_2		R_{22}	R_{23}	...	R_{2k}	...	R_{2l}

3	N_3	R_{33}	...	R_{3k}	...	R_{3l}
\vdots	\vdots			\vdots		\vdots
k	N_k			R_{kk}	...	R_{kl}

^a The recovery rate in any given year can depend on both the age of the bird and the year and variation in the chances of being found in that year, so this design is confounded unless you make very restrictive assumptions.

If only young birds in the nest are ringed, eight assumptions are necessary before survival rates can be estimated from the data (Anderson *et al.* 1985):

1. The ringed sample is representative of the population.
2. There is no loss of rings.
3. Survival is not affected by ringing.
4. The year of recovery is correctly tabulated
5. The fate of each ringed individual is independent of the fates of other ringed individuals
6. Survival rates are age-specific and do not depend on the calendar year
7. The reporting rate for ringed birds is the same for all age classes and in all years of the study
8. There are no subgroups in the population that have different survival rates

The difficulty is that many of these assumptions are not valid for birds, and the statistical methods available for estimating survival rates are very sensitive to these assumptions. The main point is that this simple, biologically appealing approach to estimating survival by ringing baby birds is unacceptable statistically, and thus should not be used.

There is an extensive literature now on estimating survival in bird populations (Brownie *et al.* 1985, Conroy *et al.* 1989, Lebreton *et al.* 1992) and a set of computer programs (see Lebreton, *et al.* 1993) are available for carrying out the tedious calculations that are necessary to allow for time- and age-varying survival rates. The general procedure is to link year to year variation in ecological factors like temperature, rainfall, or population density with individual factors like sex and age in a complex survival model so that one can determine which factors are significant in

explaining survival differences. Factors that are not significant can be dropped out of the overall model, and one searches for the best simple model to fit the data.

Survival rates estimated from a long sampling program by mark-recapture type models like the Jolly-Seber model are not independent estimates and should not be analyzed by analysis of variance or regression as though they are independent (Lebreton *et al.* 1993). This is one of the main reasons that specialized methods have been developed during the last 30 years for the analysis of survival.

15.6 TESTING FOR DIFFERENCES IN SURVIVAL RATES

Survival differences among different cohorts or between the two sexes in a population need to be analyzed statistically to see if they are significant. A whole array of statistical tests are available to answer the general question of whether the life-span of individuals in population A is longer or shorter on average than that of individuals in population B. The statistical tests needed to answer these questions have been provided from medical statistics (Lee, 1992), where they are used to decide about the effectiveness of new medical treatments, and from engineering, where they are used to test hazard or failure rates of equipment. Pyke and Thompson (1986), Pollock *et al.* (1989) and Hutchings *et al.* (1991) have summarized some of the most useful parametric and non-parametric survival tests for ecologists and given examples to illustrate their utility for analyzing survivorship curves from life tables. We will illustrate only two of these tests here.

15.6.1 Log-rank Test

Two survival curves can be compared to test the null hypothesis that they are random samples from the same underlying survivorship curve. The log-rank test is one of many that can be applied to this type of data. It has the advantage of being a simple nonparametric test and the disadvantage of having relatively low statistical power. The log-rank test is a chi-squared test which compares the observed and the expected number of deaths that occur in each time interval of study. The null hypothesis is that the two groups have the same survival pattern. We expect each of the two groups to have the same number of deaths, corrected for differences in sample size of the groups. The expected number of deaths in the first sample can

be estimated by:

$$E(d_{1j}) = \frac{d_j r_{1j}}{r_j} \quad (15.41)$$

where $E(d_{1j})$ = Expected number of deaths in group 1 in time j
 r_{1j} = Number of animals at risk in group 1 at sample time j
 d_j = Total number of deaths in both groups = $\sum_{j=1}^k (d_{1j} + d_{2j})$
 r_j = Total number at risk in both groups = $\sum_{j=1}^k (r_{1j} + r_{2j})$

The variance of the number of deaths is obtained from the equation:

$$\text{var}(d_{1j}) = \frac{r_{1j} r_{2j} d_j}{r_j^2} \quad (15.42)$$

where $\text{var}(d_{1j})$ = Variance of number of deaths in group 1 at time j

and other terms are as defined above.

Given these expectations and their variance, the log-rank test statistic is:

$$\chi^2 = \frac{\left[\sum_{j=1}^k d_{1j} - \sum_{j=1}^k E(d_{1j}) \right]^2}{\sum_{j=1}^k \text{var}(d_{1j})} \quad (15.43)$$

where d_{1j} = Number of deaths in group 1 in time period j
 $E(d_{1j})$ = Expected number of deaths in group 1 in time j (eq. 15.41)
 $\text{var}(d_{1j})$ = Variance of no. of deaths in group 1 at time j (eq. 15.42)

The log-rank chi-squared test has one degree of freedom and at the usual $\alpha = 5\%$ level, the critical value is 3.84. Box 15.6 illustrates the use of the log rank test for a staggered entry design.

Box 15.6 COMPARING SURVIVAL DISTRIBUTIONS WITH THE LOG-RANK TEST

Bobwhite quail were radio-tagged in North Carolina in the autumn of 1985 and again in the autumn of 1986. More deaths occurred in 1986 and the wildlife biologist wishes to know if the survival patterns shown in these two years differ significantly.

First, combine the data for the two samples:

Week	Fall 1985		Fall 1986		Combined data	
	No. at risk r_{1j}	No. of deaths, d_{1j}	No. at risk r_{2j}	No. of deaths, d_{2j}	No. at risk r_j	No. of deaths, d_j
1	7	1	7	0	14	1
2	6	0	6	0	12	0
3	8	0	11	1	19	1
4	13	0	10	0	23	0
5	18	0	16	1	34	1
6	18	0	15	0	33	0
7	18	0	15	1	33	1
8	18	0	14	0	32	0
9	18	0	14	3	32	3
Total		1		6		7

Then proceed as follows to compute the log rank test statistic:

1. Calculate for each of the time periods in the study the number of deaths expected under the null hypothesis that the two samples have the same survival pattern. From equation (15.41):

$$E(d_{1j}) = \frac{d_j r_{1j}}{r_j}$$

where d_{1j} = Number of deaths occurring in sample 1 at time j
 r_{1j} = Number of animals at risk in sample 1 at time j

$$E(d_{11}) = \frac{1(7)}{14} = 0.5$$

$$E(d_{12}) = \frac{0(6)}{12} = 0$$

$$E(d_{13}) = \frac{1(8)}{19} = 0.421$$

and the resulting values are given in the table below.

2. Calculate the variance of the number of deaths from equation (15.42). Note that if no deaths are observed in a time period, there is no variance (as in week 2):

$$\text{var}(d_{1j}) = \frac{r_{1j} r_{2j} d_j}{r_j^2}$$

$$\text{var}(d_{11}) = \frac{7(7)(1)}{14^2} = 0.25$$

$$\text{var}(d_{13}) = \frac{8(11)(1)}{19^2} = 0.244$$

Expected no. of deaths in first sample	Variance of expected no. of deaths
0.500	0.250
0	0
0.421	0.244
0	0
0.529	0.249
0	0
0.545	0.248
0	0
1.687	0.738

Total 3.683 1.729

3. Calculate the chi-squared test statistic for the log-rank test from equation (15.43):

$$\chi^2 = \frac{\left[\sum_{j=1}^k d_{1j} - \sum_{j=1}^k E(d_{1j}) \right]^2}{\sum_{j=1}^k \text{var}(d_{1j})} = \frac{[1 - 3.683]^2}{1.729} = 4.16$$

This chi-square statistic tests the null hypothesis that the two survival distributions are the same and has one degree of freedom. At $\alpha = 5\%$, the critical value of χ^2 is 3.84, and thus we reject the null hypothesis for these quail survival data. Survival was significantly higher in fall 1985.

These calculations can be carried out in Program LOG RANK (Appendix 2, page 000)

The log rank test assumes that survival times are independent for the different animals. If you have banding or tagging data, it assumes that the tags or bands do not reduce survival. If you are dealing with radio-telemetry data, it assumes that carrying a radio-tag does not reduce survival. Censored data in radio-telemetry or mark-recapture studies represent a particular problem for survival estimation, since we assume that censored individuals go on living a normal life span. But it is possible that a radio failure is caused by a predator destroying the radio while killing the individual. In this case an error would be introduced and the survival rate would be lower than estimated. Upper and lower limits of errors from censoring can be estimated by assuming that every censored individual was in reality a death (to give a lower bound) and by assuming that every censored animal lived until the end of the study period (to give an upper bound).

15.6.2 Likelihood Ratio Test

The log-rank test is a non-parametric test, and in some cases it is more useful to use a parametric test. Parametric tests for survival differences assume a specific survival distribution model, like the exponential distribution shown in Figure 15.1. If survival distributions for two groups are the same, and less than 20% of the observations are censored, and sample size is large ($n_1 + n_2 > 50$) then the likelihood ratio test has greater power than the log rank test (Lee 1992). It is important to check that the survival distribution is the same for the two groups being compared before using this test.

The likelihood ratio test proceeds in six steps:

1. Sum the censored and uncensored life-spans of individuals in the two treatments:

$$\begin{aligned} x_i &= \sum_{j=2}^k (c_{ij} j) + s_i k \\ y_i &= \sum_{j=2}^k d_{ij} j \end{aligned} \tag{15.44}$$

where

x_i = Sum of censored life spans for group i ($i = 1, 2$)

y_i = Sum of uncensored life spans for group i

c_{ij} = Number of censored individuals in group i at time period j

d_{ij} = Number of deaths in group i at time period j

s_i = Number of survivors to the end of the study for group i ($i = 1, 2$)

k = Last time period of the study

2. Compute the maximum likelihood estimates of the death rate per time interval:

$$\hat{\lambda}_i = \frac{d_i}{x_i + y_i} \quad (15.45)$$

where λ_i = Finite death rate for group i per unit time for the entire study

and the other terms are defined above.

3. Calculate the combined individual likelihood functions for the two groups under the assumption that they have significantly different survival rates:

$$L(\hat{\lambda}_1, \hat{\lambda}_2) = \lambda_1^{d_1} \lambda_2^{d_2} \left(e^{-\hat{\lambda}_1(x_1+y_1) - \hat{\lambda}_2(x_2+y_2)} \right) \quad (15.46)$$

4. Repeat these calculations of the death rate for the entire sample combined under the null hypothesis of equal survival in the two groups:

$$\hat{\lambda} = \frac{d_1 + d_2}{x_1 + y_1 + x_2 + y_2} \quad (15.47)$$

5. Repeat the calculation of the likelihood function for the combined sample under the null hypothesis of equal survival:

$$L(\hat{\lambda}, \hat{\lambda}) = \hat{\lambda}^{d_1+d_2} \left(e^{-\hat{\lambda}(x_1+y_1+x_2+y_2)} \right) \quad (15.48)$$

6. Calculate the test statistic for the likelihood ratio as follows:

$$\chi^2 = -2 \log_e \left(\frac{L(\hat{\lambda}, \hat{\lambda})}{L(\hat{\lambda}_1, \hat{\lambda}_2)} \right) \quad (15.49)$$

This chi-square value for the likelihood ratio test has one degree of freedom and tests the null hypothesis of equal survival distributions in the two groups.

Program SURVIVAL (Appendix 2, page 000) can calculate the log rank test and the likelihood ratio test for the comparison of survival in two groups.

15.6.3 Temporal Differences in Mortality Rates

Once you have determined that your two groups differ in mortality rates, it may be useful to try to pinpoint the exact time or age at which the differences occur. The simplest approach to this problem is to compute confidence limits for the time-specific mortality rates (Hutchings *et al.*, 1991).

$$\hat{m}_j = \frac{d_j}{r_j - 0.5(d_j)} \quad (15.50)$$

where \hat{m}_j = Finite rate of mortality during the time interval j to $j+1$
 d_j = Number of individuals dying during the time interval j to $j+1$
 r_j = Number of individuals alive at start of time interval j

The variance of this mortality rate is given by:

$$\text{var}(\hat{m}_j) = \frac{(\hat{m}_j)^2}{d_j} \left\{ 1 - \left(\frac{\hat{m}_j}{2} \right)^2 \right\} \quad (15.51)$$

This variance can be used to estimate confidence limits around the mortality rates for each time period in the usual manner (e.g. equation 15.42). Precision of survival rates will be poor unless the sample size is 20 or more in each time period.

15.7 SUMMARY

Survival rates can be calculated in many different ways and this chapter summarizes some of the more common methods of analysis. Survival rates can be expressed as finite or as instantaneous rates. Finite rates are easier to comprehend but instantaneous rates are more convenient for mathematical manipulations.

Life tables are a convenient method of summarizing the age-specific mortality schedule of a population. Six methods have been used for constructing life tables and each makes specific assumptions about the population. Key-factor analysis can utilize life tables gathered over several years to determine the best way to predict population changes. It is useful in pointing to the age or stage in the life cycle that is critical for causing changes in numbers.

Survival rates can be estimated from age composition data, if survival rates are constant. Fisheries managers have used age composition data most effectively to estimate the survival components of commercially-harvested species.

Radio-telemetry data can be used to estimate survival rates with high precision if enough individuals are followed closely. The estimation problem is identical to that for estimating nesting success in birds from periodic visits to nests. Survival rate estimation from mark-recapture kinds of data are more difficult to estimate because not all individuals are recaptured or found dead. Maximum likelihood estimators have been developed and computer programs are available to do the tedious calculations necessary to answer demographic questions about variation in survival rates in space and time among different sex and age groups in the population.

Statistical comparisons of survival between two groups can be done with a variety of parametric and non-parametric tests. The commonly used log-rank test has low statistical power unless samples are large. There is no substitute for large ($n > 40$) samples if you wish to have good precision in survival studies.

SELECTED READINGS

- Bart, J. & Robson, D.S. 1982. Estimating survivorship when the subjects are visited periodically. *Ecology* **63**: 1078-1090.
- Caughley, G. 1977. *Analysis of Vertebrate Populations*. Chapter 8, Mortality. pp. 85-106. John Wiley and Sons, London.
- Heisey, D.M. & Fuller, T.K. 1985. Evaluation of survival and cause-specific mortality rates using telemetry data. *Journal of Wildlife Management* **49**: 668-674.
- Hutchings, M.J., Booth, K.D., & Waite, S. 1991. Comparison of survivorship by the logrank test: criticisms and alternatives. *Ecology* **72**: 2290-2293.
- Lachin, J.M. & Foulkes, M.A. 1986. Evaluation of sample size and power for analyses of survival with allowance for nonuniform patient entry, losses to follow-up, noncompliance, and stratification. *Biometrics* **42**: 507-519.
- Lebreton, J.D. & North, P.M. (eds.). 1993. *Marked Individuals in the Study of Bird Population*. Birkhäuser Verlag, Basel.
- Lebreton, J.D., Pradel, R. & Clobert, J. 1993. The statistical analysis of survival in animal populations. *Trends in Ecology and Evolution* **8**: 91-94.

Podoler, H. & Rogers, D. 1975. A new method for the identification of key factors from life-table data. *Journal of Animal Ecology* **44**: 85-115.

Pollock, K. H., Winterstein, S.R., Bunck, C.M. & Curtis, P.D. 1989. Survival analysis in telemetry studies: the staggered entry design. *Journal of Wildlife Management* **53**: 7-15.

Pyke, D.A. & Thompson, J.N. 1986. Statistical analysis of survival and removal rate experiments. *Ecology* **67**: 240-245.

QUESTIONS AND PROBLEMS

15.1. Keith and Windberg (1978) estimated the survival rate of young snowshoe hares for 265 days as:

1966-67	0.557
1967-68	0.430
1968-69	0.241
1969-70	0.339

Express these as annual finite rates of survival.

15.2 Geist (1971) gave the following data for age at death from skulls of bighorn rams (*Ovis canadensis*) from the Palliser Range, Banff National Park. Note that age 1 rams are between 1 and 2 years of age, etc.

Age at death (years)																		
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	Σ
No. found	0	2	1	3	4	0	6	6	8	9	8	13	9	3	3	1	1	77

(a) What method of life table do these data represent? Calculate a life table for these rams on the assumption that (1) $r = 0$ and (2) $r = -0.02$. How do they differ?

(b) Calculate the mean expectation of life for an individual born into this population.

15.3 Smith and Polacheck (1981) reported the following age composition data for female northern fur seals (*Callorhinus ursinus*) from 1958-1961:

Age (years)	1958	1959	1960	1961
3	39	43	18	84
4	42	93	36	96
5	70	114	55	68

6	99	118	45	62
7	103	143	66	95
8	102	164	105	107
9	81	108	144	114
10	97	96	129	112
11	113	98	136	82
12	134	76	106	71
13	110	56	120	76
14	92	70	107	67
15	71	87	67	68
16	56	69	53	55
17	36	36	46	24
18	22	27	23	25
19	14	16	19	10
20-22	5	17	12	9

- (a) Can you calculate a life table from these data? Why or why not?
- (b) Can you estimate survival rates from these data? Are survival rates constant with age?

15.4 Smith (1987) gives the age structure of the ringed seal (*Phoca hispida*) from the 1980 and 1981 catch at Holman Island in the western Canadian arctic as follows:

Age	1980	1981
0+	195	118
1	61	87
2	26	89
3	59	82
4	13	73
5	32	57
6	21	52
7	8	15
8	8	10
9	28	16
10	24	17
11	19	11
12	10	13
13	16	11
14	15	15
15	10	16
16	6	6
17	4	7
18	2	4
19	1	7

20	3	4
Totals	561	710

Calculate an average annual survival rate from these catch curves. What assumptions must be made to make these estimates?

- 15.5** Mayfield (1961) reported the following data on nest loss for Kirkland's warbler: 154 nests located, total exposure = 882.5 nest days, with a total of 35 nests lost during incubation. Estimate the probability of a nest surviving the 14-day incubation period for this species. How large a sample would be needed to keep the confidence limits to ± 0.02 (rate per day)?
- 15.6** Twenty moose were radiocollared and followed through a winter (October-March). Two individuals died, one on November 17 and one on January 6. Two radiocollars failed; no signal was received after December 27 on one and February 22 on the other. Estimate the survival rate for the six-month period.
- 15.7** How much precision would be lost by doing radio-telemetry checks every second day instead of every day? every 5th day? On what factors will this loss of precision depend? Formalize your intuition on this question and then read Bart and Robson (1982, *Ecology* 63: 1085) for a discussion of this problem.
- 15.8** Waterfowl biologists have often estimated the success of duck nests by dividing the number of nests from which eggs hatch by the number of nests found (Johnson and Klett 1985). Discuss whether this procedure will give an unbiased estimate of nest success and under what conditions.
- 15.9** Mark O'Donoghue radio-collared Canada lynx in the Yukon during the winter of 1990-91 and obtained these results:

Week	No. at risk	No. deaths	No. censored	Week	No. at risk	No. deaths	No. censored
1	8	0	0	16	16	0	0
2	8	0	0	17	16	0	0
3	8	0	0	18	16	0	0
4	8	0	0	19	16	0	0
5	12	0	0	20	16	1	1
6	12	0	0	21	14	0	0
7	14	0	0	22	14	0	1
8	14	1	0	23	13	0	0
9	13	2	0	24	13	0	0
10	12	0	0	25	15	0	0
11	14	0	0	26	15	0	0
12	16	0	0	27	15	0	5
13	17	0	0	28	10	0	0
14	17	0	0	29	10	0	0
15	17	1	0	30	10	0	0

- (a) Estimate the overwinter survival rate for lynx during this time period.
 - (b) Convert this survival estimate to an annual rate of survival.
 - (c) What difference would it make to your conclusions if the 5 lynx censored in week 27 had died at the same time their radios stopped functioning?
 - (d) Suggest two ways of improving the precision of the survival estimate.
 - (e) What would be gained in this study by checking the lynx daily instead of weekly?
 - (f) What is wrong with calculating the survival rate by noting that there were 22 lynx studied in total over the winter and 5 deaths were observed, so the mortality estimate is $5/22 = 23\%$ (77% survival rate)?
- 15.10** Recompute the expectation of further life data in Box 15.4 (page 000) using equation (15.17) as a better estimate of L_x . Discuss the magnitude of the difference in the two estimates of e_0 and which estimate is likely to be more accurate for these data.