

18

KEY FACTOR ANALYSIS

In collaboration with David Bonter

Objectives

- Simulate a population that has nonoverlapping generations.
- Use the beta distribution.
- Calculate the stage-specific mortality, K_x , for each stage in the life cycles.
- Conduct a key stage analysis of the various stages in the life cycle.

Suggested Preliminary Exercise: Life Tables, Survivorship Curves, and Population Growth

INTRODUCTION

Let's assume you've been tracking the population dynamics of an annual plant through its life cycle. You tediously count the number of seeds the plant sets, then count the number of seedlings to estimate the germination rate, then count the number of vegetative rosettes, the number of flowering adults, then the number of fruiting adults. Thus you have tracked the fates of individuals in one stage and counted how many individuals survived to the next stage. If this was an endangered plant, you might want to know the stage of the life cycle in which the highest mortality occurs. For example, you might find out that the total mortality across the life cycle is strongly influenced by the failure of seeds to germinate, or by the failure of flowering plants to produce fruit. With such information, you can potentially target your efforts to reducing mortality at that particular stage.

The attempt to identify factors responsible for population change and to assess the magnitude of their effects is called **key factor analysis**. This analysis was developed by Morris (1959) to study spruce budworm outbreaks in forests in eastern Canada. Key factor analysis is specifically for organisms with discrete (nonoverlapping) generations, in which a single age class is present at any given time. The analysis, for example, could be applied to an insect population that moves from egg to larval to pupae to adult stages. The method also assumes that a series of different mortality factors operate on the population sequentially. For example, if two parasites and one disease kill larval insects, key factor analysis assumes that parasite A acts first to kill a sample, then parasite B kills a portion, then disease C acts to kill some of the remaining individuals (Krebs 1999: 511).

Modeling Key Factors

To set up a spreadsheet model of key factor analysis, we will let

- N_x denote the number of individuals alive at any given stage.
- N_{x+1} denote the number of individuals at the next stage.
- b_x denote the per capita birth rate of reproducing adults.
- k denote the stage-specific mortality, or “killing power.”
- K denote the total generational mortality, or the sum of all the k 's.

The main idea behind key factor analysis is that by comparing N_x in one stage to N_x in the previous stage, we can identify which stage has the largest mortality. We can also add up the k 's to calculate K , the total mortality across generations. The k factors indicate the importance of a particular stage to the total generational mortality, and the k factor that most strongly affects generational mortality, K , is called the **key factor**.

The steps in a key factor analysis (Varley and Gradwell 1960) include:

1. Computing the observed fecundity, which is the per capita birth rate times the number of females in the population
2. Computing the population size for each stage, or N_x in a life table
3. Computing the absolute losses of individuals from one stage to the next. For stage x , the losses are computed as

$$N_x - N_{x+1}$$

4. Converting the absolute losses of individuals from one stage to the next into proportional losses. This is accomplished by taking the log of N_x . Age- or stage-specific mortality, then, is calculated as

$$\log(N_x) - \log(N_{x+1})$$

5. Defining age-specific mortality, k_x , as

$$k_x = \log(N_x) - \log(N_{x+1}) \quad \text{Equation 1}$$

6. Computing total generational loss, K , as

$$K = k_0 + k_1 + k_2 + k_3 + \cdots + k_x. \quad \text{Equation 2}$$

This analysis is done over several generations, where each generation consists of a complete life cycle and where the life cycles from one generation to the next do not overlap with each other. Each generation that is studied is a “replicate” of the key factor analysis, and these replicates are important because they let you know if a certain factor is normally a key factor, or if it is a key factor in some conditions or years but not in others.

As an example, suppose that k 's and K were computed for an insect population for 10 generations. Figure 1 shows that K , the total generational mortality, fluctuates from generation to generation. The stage-specific mortalities (small k s) are also plotted for each year and reveal the losses that occur within a stage for a single generation. The little k that most closely mimics K over time is the key factor. In this case, graphed in Figure 1, the pupal stage is the key stage. Note that a pattern could not be detected if only a single generation were studied. Which k factor is most closely tied to K can be hard to discern, especially if the k factors have similar values. In this case, the key factor can be identified by plotting the k factor against K for every single k ; Figure 2 does this for egg-stage mortality vs. total mortality.

Problems with Key Factor Analysis

You probably know by now that populations change over time through birth, death, immigration, and emigration. In fact, the equation for population growth given in Exercise 7,

$$N_{t+1} = N_t + B + I - D - E$$

is the basis for many exercises in this book. But because key factor analysis focuses on losses to a population, only death and emigration are properly represented by k fac-

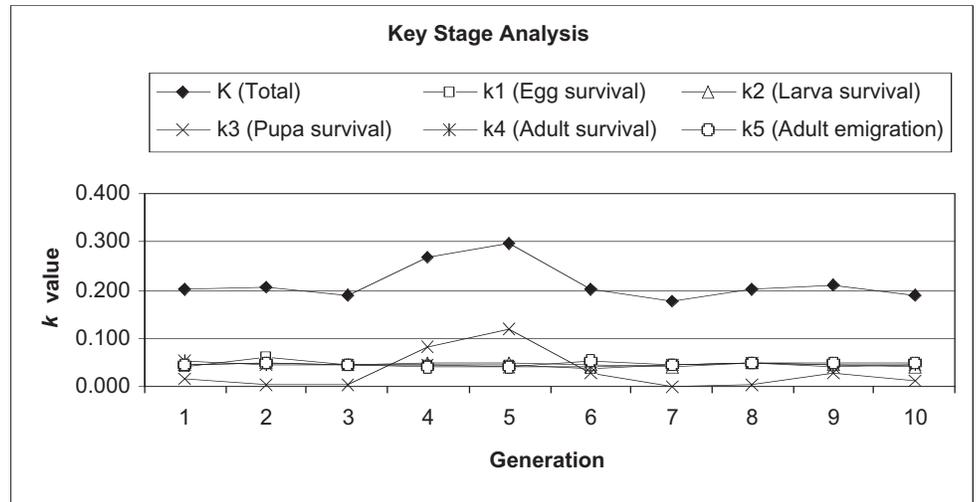


Figure 1 The total generational mortality, K , for each year is the sum of all the stage-specific mortalities for that year. The stage-specific mortalities (k_1 – k_5) are also plotted for each year, and reveal the losses that occur within a stage for a given year. The k that most closely mimics K over time is the key factor. In this case, the pupae stage is the key stage. Note that a pattern could not be detected if only a single generation was studied.

tors. The analysis also does not specifically identify factors that are responsible for population change, only the stages that are correlated with total generational mortality. The analysis gives no indication of what might be *causing* such mortality, only the stage in which it occurs. For this reason, the analysis may be more properly named *key stage analysis*. Additionally, the assumptions of key factor analysis are often violated, and many ecologists have criticized the use of traditional key factor analysis (e.g., Royama

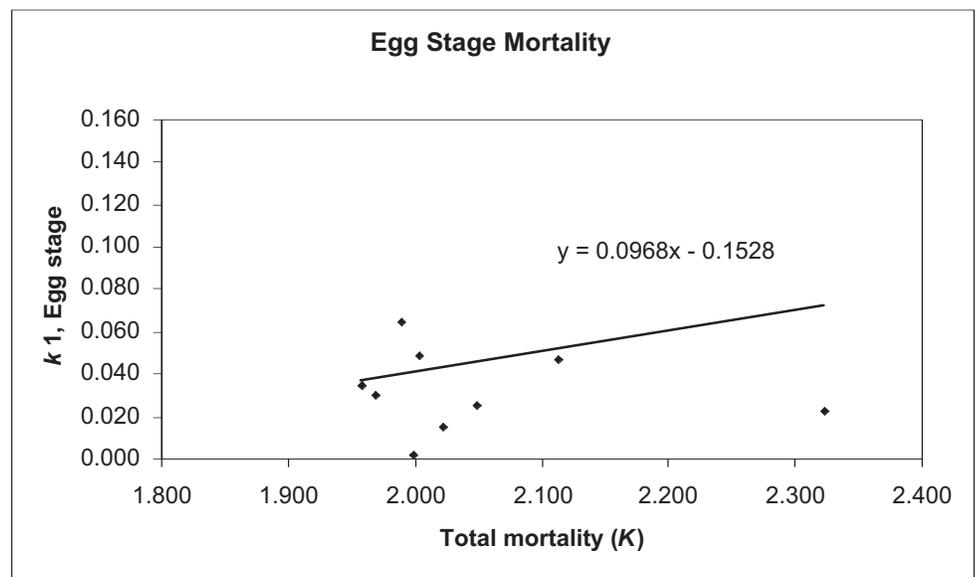


Figure 2 The relationship between k for the egg stage and total generational mortality, K , for 10 years. The slope of the regression equation is +0.0968. Similar graphs can be constructed for the other stages, and the slopes can then be compared. Were we to construct similar graphs for K and each of the other four k 's in Figure 1, the k factor that generates the highest slope with K would be the key factor.

1996). However, the traditional analysis is often used as the first step in the analysis of census data from natural populations, and several new methods have been developed that improve on the method presented here (e.g., Brown et al. 1993; Sibly and Smith 1998).

The Beta Distribution

In this exercise, we will use the beta distribution to assign probabilities that an individual will move from one stage to the next. This distribution is not used in other exercises, and we will describe it only briefly here. All probabilities range between 0 and 1, and the beta distribution (rather than the normal distribution, which can take on values greater than 1 and less than 0) is much more appropriate for modeling probabilities. The exact shape and scale of the beta distribution is controlled by two parameters, called α and β . Because you are (by now) very familiar with the normal distribution, we will take some parameters from a normal distribution that you are familiar with (μ and σ^2), and convert them into parameters from the beta distribution, α and β . For example, if survivorship is known to have a mean, \bar{x} , of 0.6, and a standard deviation, S , of 0.1, this corresponds to $\alpha = 13.8$ and $\beta = 9.2$. A beta distribution with these parameters will show that most probabilities are 0.6, but there is substantial variation from sample to sample. The values of α and β can be calculated as follows, where the sample mean and standard deviation, \bar{x} and S^2 , estimate μ and σ^2 :

$$\alpha = \frac{\beta * \mu}{1 - \mu} \qquad \beta = \mu - 1 + \frac{\mu * (1 - \mu)^2}{\sigma^2}$$

In this way, we can include variation in survival probabilities with an appropriate distribution (the beta distribution). However, you can intuitively visualize the probabilities based on your experience from working with normal distributions. (Thanks to Jeff Buzas at the University of Vermont, who provided these conversions). Figure 3 shows how the conversion works for a mean survival = 0.6 and a standard deviation = 0.1. These parameters translate into a beta distribution whose $\alpha = 13.8$ and $\beta = 9.2$. If we changed α and β in Figure 3, the distribution would take on a new shape.

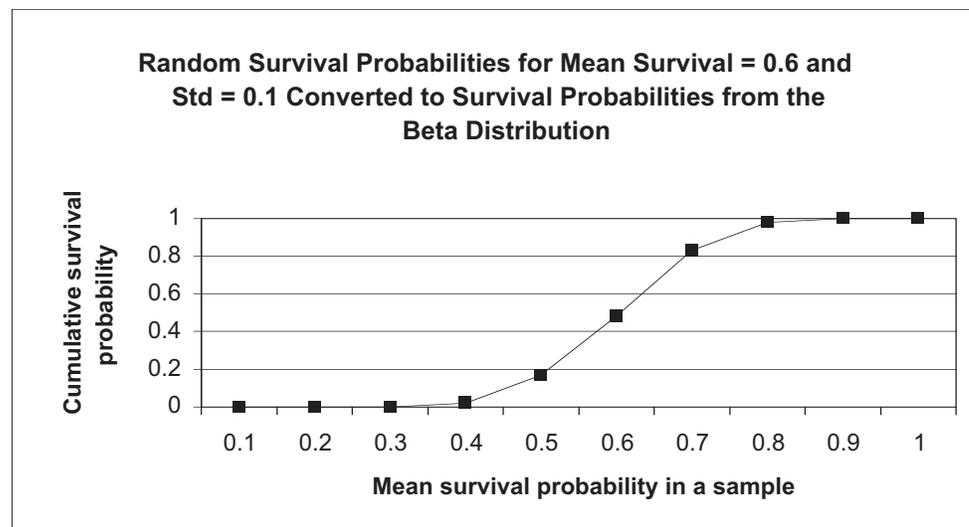


Figure 3 The x -axis gives the survival probability for a single sample. The cumulative probabilities are given on the y -axis. Look at the x -axis and note where the cumulative survival probabilities change very little. You should see that they change very little for probabilities $P < 0.4$ and $P > 0.8$. In between these values, the cumulative survival probabilities increase dramatically, suggesting that most of the data points in this distribution fall between 0.5 and 0.7. At $P = 0.6$, the cumulative probability equals 0.5, suggesting that half of the observations in the data set fall above 0.6 and half fall below 0.6, as expected.

PROCEDURES

In this exercise, you'll model a hypothetical insect population that moves through several stages in its life cycle. We'll assume the population is large and that we can track the total number of individuals alive at each stage. We'll assign probabilities that individuals move from one stage to the next, and then calculate the k factors and identify the key mortality factor. You'll assign probabilities that individuals move from one stage to the next with the beta distribution, and then calculate the k factors and identify the key mortality factor.

As always, save your work frequently to disk.

INSTRUCTIONS

A. Set up the model population.

1. Open a new spreadsheet and set up headings as shown in Figure 4.

2. Enter parameter estimates (means and standard deviations) in cells B5–G6.

3. Draw random values from a beta distribution for survival probability at each stage.

4. In cells C8–G8 enter a formula for the β parameter of a beta distribution. In cells C7–G7, enter a formula that will calculate the α parameter.

ANNOTATION

	A	B	C	D	E	F	G	H
1	Key Factor Analysis							
2		Population variables						
3		Eggs laid	Egg survival	Larva survival	Pupa survival	Adult survival	Adult fidelity	Number of females
4								
5	Mean =	100	0.90	0.40	0.60	0.80	0.95	1000.00
6	Standard deviation =	30.00	0.10	0.05	0.10	0.10	0.01	
7	Beta distribution α =							
8	Beta distribution β =							

Figure 4

Enter the values shown in Figure 4 in cells B5–G6.

We'll consider an insect with nonoverlapping generations and whose life cycle consists of a series of mortality factors that operate in a linear sequence with no interaction. Eggs are laid by adults and hatch with some probability, then move to the larvae stage, and pupate to become adults. The probability of moving from one stage to the next is defined by a probability between 0 and 1. Some adults are capable of moving away from the study area population (emigration). The probability of remaining in the population and not emigrating is given in the column labeled "Adult fidelity."

We'll add an element of stochasticity to our model by establishing means and variances for each parameter, and then "drawing" a random number from these distributions. In previous exercises, you may have used the **NORMINV** function to draw a random probability from a normal distribution with a given mean and standard deviation. The spreadsheet then converts this probability into a data point. This function won't work for survival probabilities, though, because our survival probabilities can only take on values between 0 and 1. For survival probabilities, the distribution we must draw at random from a beta distribution. The parameters in the beta distribution are α and β (made by typing "a" or "b" in on your keypad and then changing the font to the Symbol font).

Although α and β are not the same thing as means and standard deviations, we can enter these formula based on the conversion equations

$$\alpha = \frac{\beta * \mu}{1 - \mu} \quad \beta = \mu - 1 + \frac{\mu * (1 - \mu)^2}{\sigma^2}$$

In cell C8, enter the formula =C5-1+((C5*(1-C5)^2)/C6^2) . Copy the formula across to cell G8.

In cell C7 enter the formula =(C8*C5)/(1-C5). Copy the formula across to cell G7.

5. Your spreadsheet should now resemble Figure 5. Save your work!

Now α and β are mathematical functions of the means and standard deviations specified in rows 5 and 6. As a result, we can draw random probabilities between 0 and 1 that have the means and standard deviations we specify.

	A	B	C	D	E	F	G	H
1	Key Factor Analysis							
2		Population variables						
3		Eggs laid	Egg survival	Larva survival	Pupa survival	Adult survival	Adult fidelity	Number of females
4								
5	Mean =	100	0.90	0.40	0.60	0.80	0.95	1000.00
6	Standard deviation =	30.00	0.10	0.05	0.10	0.10	0.01	
7	Beta distribution =		7.20	38.00	13.80	12.00	450.30	
8	Beta distribution =		0.80	57.00	9.20	3.00	23.70	

Figure 5

B. Determine model inputs for Years 1-10.

1. Set up new headings as shown in Figure 6, but extend years to year 10 in cell K10.

	A	B	C	D	E
10	Model Inputs	Year 1	Year 2	Year 3	Year 4
11	Eggs laid				
12	Egg survival				
13	Larva survival				
14	Pupa survival				
15	Adult survival				
16	Adult fidelity				

Figure 6

2. In cells B11–K11, enter a formula to give the mean number of eggs laid in year 1.

In cell B11 enter the formula `=NORMINV(RAND(),B5,B6)`. Copy this formula over to cell K11.

Here we do use the **NORMINV** function. The **NORMINV** function returns the *inverse* of the normal cumulative distribution, given a mean and standard deviation. It has the syntax **NORMINV(probability,mean,standard_dev)**. The B11 formula draws a random probability from a distribution whose mean is given in cell B5 and whose standard deviation is given in cell B6. The spreadsheet then converts this probability into an actual data point from the distribution, which is the number of eggs laid in year 1. Note that when you press F9, the calculate key, the spreadsheet will generate a new random number, which means that a new random number is drawn from the distribution and hence a new average fecundity computed.

3. In cells B12–K12, enter a formula to give the probability that eggs will hatch.

In cell B12 enter the formula `=BETAINV(RAND(),C7,C8)`. Copy the formula over to cell K12.

The B12 formula gives the probability that eggs will hatch. Since this is a probability whose values must fall between 0 and 1, we use the beta distribution (instead of the normal distribution). The **BETAINV** formula functions like the **NORMINV** formula, except that the distribution is a beta distribution instead of a normal distribution. The formula in cell B12 tells the spreadsheet to draw a random cumulative probability from the beta distribution whose parameters are α (cell C7) and β (cell C8). (Remember, you entered formulae to compute α and β based on the means and standard deviations entered in rows 5 and 6.) The spreadsheet converts the cumulative probability into a data point, which is the probability that eggs will hatch in year 1. Press F9 to generate a new estimate.

4. Enter formulae in cells B13–B16 to determine random probabilities, drawn from the beta distribution. Copy your formulae across to column K.

5. Double-check results.

6. Save your work.

We used the following formulae:

- Cell B13 =**BETAINV(RAND(),\$D\$7,\$D\$8)**
- Cell B14 =**BETAINV(RAND(),\$E\$7,\$E\$8)**
- Cell B15 =**BETAINV(RAND(),\$F\$7,\$F\$8)**
- Cell B16 =**BETAINV(RAND(),\$G\$7,\$G\$8)**

Your spreadsheet should now resemble Figure 7, although your numbers will probably be different due to the random sampling from the normal and beta distributions.

	A	B	C	D	E
10	Model Inputs	Year 1	Year 2	Year 3	Year 4
11	Eggs laid	96.9	151.5	72.1	119.5
12	Egg survival	0.9	1.0	0.9	0.9
13	Larva survival	0.4	0.3	0.4	0.4
14	Pupa survival	0.6	0.7	0.7	0.6
15	Adult survival	0.9	0.8	0.7	0.6
16	Adult fidelity	1.0	0.9	0.9	0.9

Figure 7

C. Calculate model outputs and project growth for 10 years.

1. Set up new spreadsheet headings as shown in Figure 8, but extend your years to year 10.

	A	B	C	D	E
18	Model outputs (Nx)	Year 1	Year 2	Year 3	Year 4
19	Eggs				
20	Eggs hatched				
21	Larvae				
22	Pupae				
23	Adults				
24	Nonemigrants				

Figure 8

2. In cell B19, enter a formula to calculate the actual number of eggs laid in year 1.

In cell B19 enter the formula =**\$H\$5*B11**.

The actual number of eggs laid is the average fecundity \times the number of females.

3. In cell B20, enter a formula to calculate the number of eggs hatched in year 1.

In cell B20 enter the formula =**B12*B19**.

The number of eggs hatched is a function of hatching probability calculated in cell B12.

4. In cells B21–B24, enter formulae to compute numbers of individuals in various stages. Copy each formula across to column K to complete 10-year simulation.

We used the following formulae:

- Cell B21 =**B20*B13**
- Cell B22 =**B21*B14**
- Cell B23 =**B22*B15**
- Cell B24 =**B23*B16**

D. Set up the *k* factor analysis.

1. Set up new headings as shown in Figure 9, but extend your years to year 10.

2. In cell B28, enter a formula to calculate the mortality due to number of eggs that failed to hatch.

3. In cells B29–B32, enter formulae to compute *k* for the remaining stages.

4. In cell B27, sum the *k* values for year 1 to generate the *K* value.

5. Compute the *k* and *K* values for years 1–10. Save your work.

E. Create graphs.

1. Graph *K* and the *k*'s as a function of time. Which *k* factor appears to “track” *K* the most?

Now we can estimate the the stage-specific mortalities—the *k* factors (“little *k*’s”)—for each stage in the life cycle.

	A	B	C	D	E
26	Key factor analysis	Year 1	Year 2	Year 3	Year 4
27	K (Total)				
28	k1 (Egg survival)				
29	k2 (Larva survival)				
30	k3 (Pupa survival)				
31	k4 (Adult survival)				
32	k5 (Adult emigration)				

Figure 9

In cell B28 enter the formula **=LOG(B19)-LOG(B20)**.

As shown in Equation 1, age-specific mortality is calculated by subtracting each log of the population size from the previous one:

$$k_x = \log(N_x) - \log(N_{x+1})$$

Thus, the formula in cell B28 gives the *k* value or the mortality due to the number of eggs that failed to hatch.

We used the following formulae:

- Cell B29 **=LOG(B20)-LOG(B21)**
- Cell B30 **=LOG(B21)-LOG(B22)**
- Cell B31 **=LOG(B22)-LOG(B23)**
- Cell B32 **=LOG(B23)-LOG(B24)**

Cell B32 does not give a mortality value *per se*, because it reflects the loss of individuals due to emigration rather than death. However, emigration has the same effect on the population as mortality in that emigrants will not contribute to the next generation.

In cell B27 enter the formula **=SUM(B28:B32)**.

Copy the formula in cell B27 across columns to column K.

Use the line graph option and label your axes fully. Your graph should resemble Figure 10.

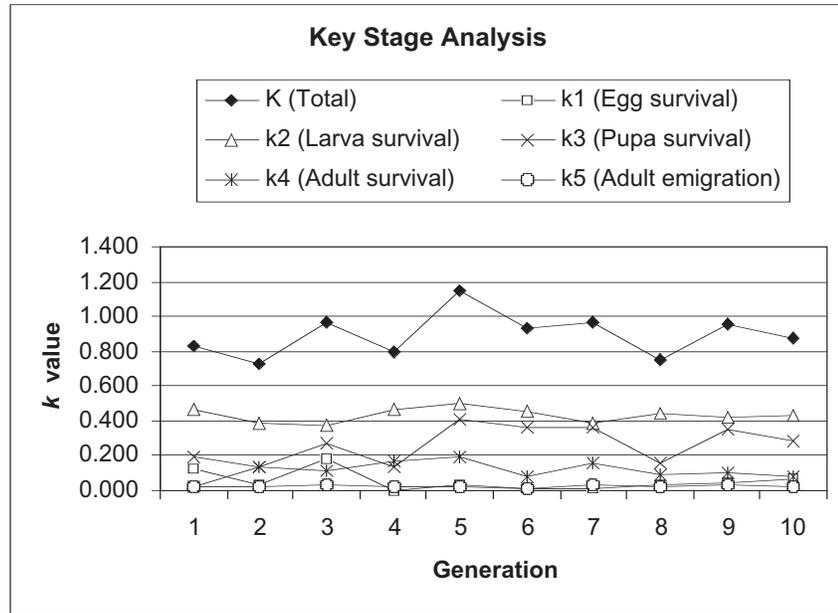


Figure 10

2. Press F9, the calculate key, to simulate new conditions over time. Does your key factor appear to change?

3. For each *k*, construct a scatter graph that plots *k* against *K*. Add trendlines (slope) for each graph.

Add trendlines by selecting the chart; then go to the Chart menu, select Add Trendline, and add a Linear trendline. Then click on Options | Display equation on chart. Your graph should look something like Figure 11.

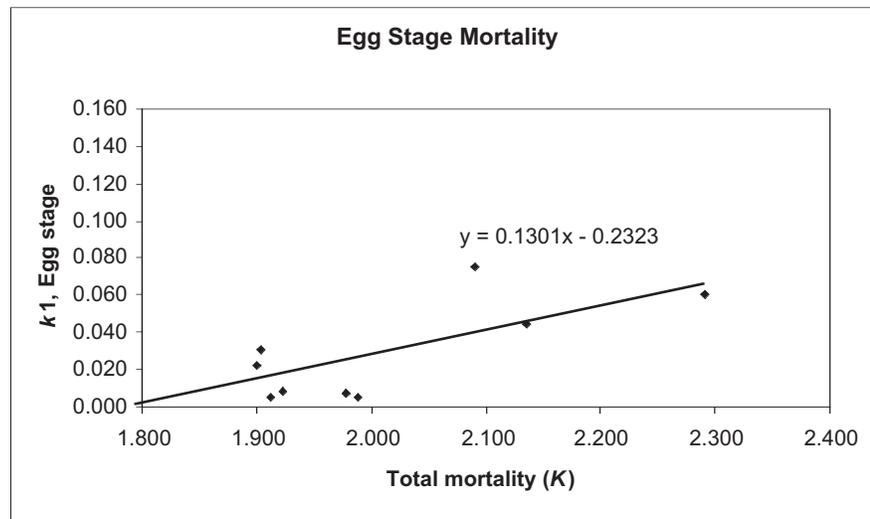


Figure 11

Compare the slopes of *K* versus *k* for each stage. The *k* value that has the greatest slope is the key factor.

4. Press F9 to generate new data. Does the key factor appear to change?

5. Answer questions 1 and 2 at the end of the exercise.

F. Conduct 100 trials.

1. Set up column headings as shown in Figure 12.

2. In cell B36, use the **SLOPE** function to compute the slope of the regression between k_1 and K .

3. In cells C36–F36, compute the slopes of other k regressions.

4. Set up a linear series from 1 to 100 in cells A37–A136.

5. Write a macro to track k versus K regression slopes and run it for 100 trials. Save your work!

6. Compute the average slope with the **AVERAGE** function to determine which k has the largest slope with K . This is the key stage.

Because the k factor appears to change from trial to trial, it would be useful to conduct 100 trials, tracking the slopes of each k versus K regression equation, and then computing the average slope for the 100 trials. This will give you a better indication of which k factor has the greatest regression slope with K . There are many ways you could do this. A suggested way follows.

	A	B	C	D	E	F
34		Regression slopes				
35	Trial	k1	k2	k3	k4	k5

Figure 12

In cell B36 enter the equation **=SLOPE(B28:K28,B27:K27)**. Your answer should match the slope displayed on your graph that is analogous to Figure 11.

- Cell C36 **=SLOPE(B29:K29,B27:K27)**
- Cell D36 **=SLOPE(B30:K30,B27:K27)**
- Cell E36 **=SLOPE(B31:K31,B27:K27)**
- Cell F36 **=SLOPE(B32:K32,B27:K27)**

Open the record macro function (see Exercise 2). Assign a shortcut key, then record the following steps:

- Press F9, the calculate key, to generate new data, and hence new slopes.
- Select cells B36–F36. Open Edit | Copy.
- Select cell B35.
- Open Edit | Find. Leave the Find What box blank and search by columns. Select Find Next, then Close. Your cursor should move down to cell B37.
- Open Edit | Paste Special and paste in the values.

Open Tools | Macro | Stop Recording. Now when you press your shortcut key 99 more times, your results (the slopes of each k versus K) will be recorded for each trial.

In cell A137 type "Average."

In cell B137 enter the equation **=AVERAGE(B37:B136)**.

Copy this equation over to cell F137.

QUESTIONS

1. Fully interpret the k factors in your figures. Which factor appears to be the key factor in your model?
2. Press F9 to generate new sets of data, and inspect your plot of k 's and K over generations. Does your key factor change with new simulations?
3. Compute the average of the regression slope estimates from your 100 trials. Which k factor has the highest regression coefficient when regressed against K ?
4. Based on the original population variables, and assuming our hypothetical insect population is endangered, did the key factor analysis assist you in developing management recommendations? If so, how?
5. Change the parameter values in cells B5–H6 so that the standard deviation of all parameters is 0.001 (little variation over generations). Clear your macro results (cells B37–F135) and run your macro again. When the parameters do not vary from generation to generation, which stage is the key factor?
6. Change the parameter values in cells B5–H6 so that all survival probabilities equal 0.7. Increase one of the standard deviations (e.g., cell D6) to 0.1. Clear your macro results (cells B37–F135) and run your macro again. When the parameters are equal but one stage is variable, which stage is the key factor?

	A	B	C	D	E	F	G	H
2		Population variables						
3		Eggs laid	Egg survival	Larva survival	Pupa survival	Adult survival	Adult fidelity	Number of females
4								
5	Mean =	100	0.70	0.70	0.70	0.70	0.70	1000.00
6	Standard deviation =	30.00	0.01	0.10	0.01	0.01	0.01	

LITERATURE CITED

- Brown, D., N. D. E. Alexander, R. W. Marris and S. Albon. 1993. Structured accounting of the variance of demographic change. *Journal of Animal Ecology* 62: 490–502.
- Krebs, C. J. 1999. *Ecological Methodology*. Addison-Wesley, New York.
- Morris, R. F. 1959. Single factor analysis in population dynamics. *Ecology* 40: 580–588.
- Royama, T. 1996. A fundamental problem in key factor analysis. *Ecology* 77: 87–93.
- Sibly, R. M. and R. H. Smith. 1998. Identifying key factors using λ contribution analysis. *Journal of Animal Ecology* 67: 17–24.
- Varly, G. C., and G. R. Gradwall. 1960. Key factors in population studies. *Journal of Animal Ecology* 29: 399–401.