

An Introductory Tutorial:
Learning R for Quantitative Thinking in the Life Sciences

Scott C Merrill

October 31st, 2012

Chapter 8

Efficiency

Much of the programming we have learned in R has followed largely traditional programming approach. The steps we have taken have been primarily stepping through logical sequences. To a true R programmer, much of our programming would be considered very chunky and ugly. One of the strengths of R is that there are many ways to approach and code a problem. The following short presentation provides an excellent example. Run through this presentation and run some of the code:

<http://www.stat.auckland.ac.nz/~stat782/downloads/10-Efficiency.pdf>

Modeling exercise (Largely from Hobbs)

One of the ways that we can estimate parameters in non-linear models is to minimize the sum of squared differences between the model predictions and the observations. This is to say that we are searching for a set of initial conditions and parameter values that minimize your error (predicted by model minus the observed).

Find the best-fit parameter values (and initial conditions) for a logistic model of growth of the Rocky Mountain National Park elk population (Data on my website under the courses tab)

$$N_{(time)} = N_{(time-1)} + r * N_{(time-1)} * \left(1 - \frac{N_{(time-1)}}{K}\right)$$

by finding the minimum of the sum of squared differences between the observed population size and the predicted population size. Use the data set of observations that you prepared as part of the R tutorial.

Here are some hints:

- 1) To make this problem a bit easier, you can reasonably guess that the value of r is somewhere between .1 and .3. (Create a loop that varies r from .1 to .3 by .01)
- 2) Moreover, you can look at the data and be sure that K is somewhere between 800 and 1200 (Create a loop that searches K values from 800 to 1200 by 10)
- 3) Assume the initial population value is 350 (N at time 1)

So, to estimate the parameters, you will calculate the sum of squared error for all possible combinations of parameter values and find the combination that gives the minimum SSE.

$$\text{SSE} = (\text{Model predicted elk density} - \text{observed elk density})^2$$

Here are some more hints: Set up a matrix with 4 columns, one each for N , r , K , and SSE (where N is the predicted elk population over time). Set up 2 nested loops to vary the r and K parameters at fairly coarse intervals). Within the innermost parameter loop, you will have a loop for time, calculating the squared error for each model prediction and observation. After you exit this loop, you can calculate sum of the squared error over all of the years in your simulation (= 41). You will need a counter embedded in this loop to index the rows of the SSE array. Once you have created the SSE array, find the row of your containing the minimum value of SSE. See the section on arrays in [Hobb's] R manual if you need refreshing on this. This row will contain your parameter estimates.

After you have your program working, you can make the intervals in your for loop finer to give your more accurate estimates of the parameters (say 1 for K (over a smaller interval).

Plot the observed vs your predicted values (e.g., figure):

