**Ecological Models and Data Homework #3 – Maximum Likelihood**

**DUE: in class, Tuesday, February 11**



Fender’s blue, *Icaricia icarioides fenderii*

Silvery blue, *Glaucophyche lygdamus*

For a recent Department of Defense-funded contract to investigate how source-sink dynamics affect population viability of butterflies, Cheryl Schultz and Norah Warchola spent several years collecting data on an endangered species of butterfly, the Fender’s blue (*Icaricia icarioides fenderii*). One part of the research involved monitoring populations by walking a fixed route every week during the flight season, and counting all the butterflies seen.

Unfortunately, the Fender’s blue butterfly looks a lot like a more common species, the silvery blue butterfly. To estimate the total number of each species, they collected a few butterflies at the end of each transect walk, for careful identification. (Butterflies were released after identification.)

For this homework, you will use basic principles of likelihood to estimate the probability that each butterfly Norah and Cheryl saw in 2013 was a Fender’s blue. You will also explore the distribution of this probability in space and time.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Year | Date | Site | # of Fender's | # of Silvery |
| 2013 | 25-Apr-13 | A | 0 | 6 |
| 2013 | 25-Apr-13 | B | 0 | 4 |
| 2013 | 25-Apr-13 | C | 0 | 12 |
| 2013 | 25-Apr-13 | D | 0 | 5 |
| 2013 | 3-May-13 | A | 2 | 6 |
| 2013 | 3-May-13 | B | 2 | 6 |
| 2013 | 3-May-13 | C | 13 | 6 |
| 2013 | 3-May-13 | D | 2 | 16 |
| 2013 | 14-May-13 | A | 8 | 3 |
| 2013 | 14-May-13 | B | 3 | 4 |
| 2013 | 14-May-13 | C | 8 | 7 |
| 2013 | 14-May-13 | D | 4 | 3 |
| 2013 | 20-May-13 | A | 6 | 1 |
| 2013 | 20-May-13 | B | 2 | 7 |
| 2013 | 20-May-13 | C | 2 | 2 |
| 2013 | 20-May-13 | D | 2 | 1 |
| 2013 | 31-May-13 | A | 5 | 0 |
| 2013 | 31-May-13 | B | 2 | 0 |
| 2013 | 31-May-13 | C | 10 | 1 |
| 2013 | 31-May-13 | D | 2 | 2 |

Norah and Cheryl’s data are shown in the table to the right.

For this exercise, you should treat each row as a single binomial trial with *N* (the number of trials) equal to the total number of butterflies captured, and *k* (the number of successes) equal to the number of Fender’s blues.

The R command for a binomial probability is

dbinom(x, size, prob, log = FALSE) where “*x*” is the number of successes (*k*), “size” is the number of trials (*N*), “*prob*” is the probability of success, and “log = FALSE” determines whether the command reports actual probabilities or ln-transformed probabilities (log = TRUE gives ln-transformed probabilities; log = FALSE gives the actual probabilities).

1. Using all the data, what is the maximum likelihood proportion of Fender’s blues butterflies? [Show enough of your work to let us know how you got the answer.]
2. Graph a log-likelihood profile for this proportion, using R.
3. Explore how this probability changes in space and time by:
   1. Calculating the maximum likelihood proportion for *each day*, and graphing how this value changes through time. In other words, make a graph with the date on the *x*-axis, and the proportion for that day on the *y*-axis.
   2. Calculating the maximum likelihood proportion for *each site*, and graphing how the proportion varies among sites. In other words, make a graph with the site on the *x*-axis and the proportion for that site on the *y*-axis.
4. Calculate the *log-likelihood* of each of the following models, given Norah and Cheryl’s data:
   1. All the data have the same proportion
   2. The proportion differs among dates (but is the same for all sites)
   3. The proportion differs among sites (but is the same for all dates)

HINT #1: These are the three “models” you fit to the data in 1, 3a, and 3b

HINT #2: You can fit models to different subsets of data, and then use the “AND” rule of probability to combine likelihoods. [Technically, if you are a theorist, the “AND” rule would apply to the probabilities (of the data, given the model) that you used to calculate the likelihood. But, as a shortcut, you can combine the likelihoods.]

HINT #3: For a given subset of data, you can obtain the maximum likelihood proportions and estimates using the glm function, e.g.,

subdat <- dplyr::filter(dat, Site == “A”)

m0 <- glm(cbind(Fenders, Silvery) ~ 1, family =

binomial(link = “identity”), data = subdat)

coef(m0)

logLik(m0)

1. Which of the models in question 4 are nested? Which are non-nested?
2. Which model do you think is best for these data? Justify your answers using logic and/or formal statistics.