PubH 7407, Spring 2009: Extra Credit Project

Brown et al. (2001) describe an extensive data set that includes the presence of certain rodent species in Portal, Arizona over time. The data were collected as part of an ongoing long-term study begun in 1977 on the growth or decline of species related to environmental factors in the southwestern U.S. desert. The site is comprised of 24 quarter-hectare plots; in 16 of the plots kangaroo rats were removed and fences erected to maintain their absence. The other 8 "control" plots were left as is so kangaroo rats were free to come and go as they please. We will limit our analyses to these 8 control sites.

The 8 control sites (indexed $i = 1, \ldots, 8$) were visited periodically to see what species were trapped over 51 time periods of 6 months (indexed $j = 1, 2, \ldots, 51$). The species considered are *Dipodomys merriami* (Merriam's kangaroo rat, identified as $m_{i,j} = 0$ if none were trapped, $m_{i,j} = 1$ if at least one was trapped), *Dipodomys ordii* (Ord's kangaroo rat, $o_{i,j} = 0, 1$), *Dipodomys spectabilis* (Banner-tailed kangaroo rat, $b_{i,j} = 0, 1$), and *Perognathus baileyi* (Bailey's pocket mouse, $p_{i,j} = 0, 1$). These four species are prevalent at this location and since they eat similar foods we expect that competition among these four would be strongest relative to other species in area. We are particularly interested in how other species affect the presence or absence of Merriam's kangaroo rat, the most prevalent of the four rodents at Portal, as well as the propensity for Merriam's kangaroo rat to "take hold" once present.

A row of data for a time point i gives the current status ($\texttt{m} = m_{i,j}, \texttt{o} = o_{i,j}, \texttt{b} = b_{i,j}, \texttt{p} = p_{i,j}$), the status 6 months earlier ($\texttt{mp} = m_{i,j-1}, \texttt{op} = o_{i,j-1}, \texttt{bp} = b_{i,j-1}, \texttt{pp} = p_{i,j-1}$), the status for Merriam's one year earlier ($\texttt{mp2} = m_{i,j-2}$), and the site j. The rows are listed consecutively $j = 1, 2, \ldots, 51$ within each site $i = 1, \ldots, 8$.

There are many ways to analyze these data; I want you to consider two.

1. Fit a marginal logistic regression model with the current value of Merriam's `m` as the outcome, taking into account correlation within a site. Try fitting both current indicators (`o`, `b`, `p`) of the other rodents and indicators lagged by 6 months (`bp`, `op`, `pp`). Carefully think about what the most appropriate correlation structure might be here, and qualitatively and quantitatively describe correlation among repeated measures at a site – you might try the QIC for fun. Do not include the lagged Merriam's variables (`mp` and `mp2`) in this analysis! You are finding out probabilistically how the presence/absence of the other rodents affects Merriam's, accounting for the correlation among the presence of Merriam's at a site through the estimation procedure.

2. Examine a simple lagged (Markov) logistic regression model in which the probability of seeing Merriam's kangaroo rat (`m`) depends on whether there is currently a captured Ord's kangaroo rat, Banner-tailed kangaroo rat, or Bailey's pocket mouse (`o`, `b`, `p`), and whether there was a Merriam's kangaroo rat trapped in the previous time period (`mp`).

   Try fitting models that include a random site effect $u_1, \ldots, u_8 \overset{iid}{\sim} N(0, \sigma^2)$ and without. A random site effect can be included in the model to account for Merriem's kangaroo rat preferring certain sites. The various species have habitat preferences and the 8 control sites vary somewhat with respect to amount of vegetation, water availability, and other factors. Formally test $H_0 : \sigma = 0$. Does the inclusion of site effects change conclusions? Also compare via AIC.

   Using the "better" model (with or without random site effects), now also include the Merriam's variable lagged by a year (`mp2`). How do the results change?

Write a concise, coherent summary of your findings. This project is worth 10 points of extra credit.