Assignment # 3, Stat 8053, Fall 2013

Reading

The reading is Chapters 11 & 12 of Faraway.

Problems

The following problems are due on Friday, October 18, 2013, in class.

1. Page 230, Problem 5. Fit using kernel smoothers, loess and splines, with various choices for smoothing parameter. Use the data in this example to examine the question of whether or not temperature is increasing over time. Beyond the suggestions in the problem, you should also consider (i) fitting parametric models, perhaps linear, quadratic and cubic fits; (ii) test interesting hypotheses (Hint: you may want to design a bootstrap to do this); (iii) whether or not the pre-1881 data should be used.

2. The data in the file

   ```
   http://www.stat.umn.edu/~sandy/courses/8053/Data/BGSgirlsLong.csv
   ```

   has longitudinal data on 10 girls born in Berkeley California in 1928 or 1929. We will be interested in modeling height (the variable `Ht_cm`) as a function of `Age`. The variable `Subject` is the subject number.

   (a) Obtain a summary of the average growth curve for girls like these; you will probably want to fit a mixed effects model for this similar to the Orthodont data in the handout. You should use cross validation to select the df or the number of knots for the b-spline basis.

   (b) People interested in growth curves are particular interested in the derivative of the regression function fit in the first part. Humans are thought to undergo a "growth spurt" when the rate of increase accelerates for a time and then abates. Explore the derivative of the growth curve you have fit, and if you can derive and perform a test for a growth spurt. (Hint: my program for computing the derivative of a b-spline basis is available at

   ```
   http://www.stat.umn.edu/~sandy/courses/8053/Data/mybs.R
   ```

   If you look at the code you should be able to figure out what it does and how to use it.)

   Also included here is code that you can used to get fitted values from an lmer fit:

   ```
   lmer.predict <- function(mod){
       lin.pred <- model.matrix(mod) %*% fixef(mod)
       call.args <- names(mod@call)
       family <- if(is.na(call.args["family"]))
                   gaussian() else call.args["family"]
       family$linkinv(lin.pred)
       }
   ```

3. Shoreline development has been associated with declines in the quality of fish habitat, either through direct removal of aquatic vegetation (macrophytes) by homeowners or indirect impacts through increased runoff, shading by docks, or physical disturbance. The purpose of this analysis was to evaluate the effects of shoreline development on macrophyte species richness using dock density as an indicator of development. The Minnesota Department of Natural Resources conducts routine macrophyte surveys using transects that are placed at equidistant intervals along the shoreline with transect density proportional to lake size. Macrophytes species are identified at each transect and a species list is compiled for each lake. The data describe results of macrophytes surveys for 1,444 lakes, in addition to information on dock density and other relevant variables. Generalized linear models were used to evaluate the effect of dock density on species richness using a Poisson distribution and on offset variable when appropriate. Models were also developed to determine if dock density affected the species richness of different growth forms of macrophytes, specifically submersed species, emergent and floating-leaf species, and sensitive species. Because species richness is

affected by other natural and anthropogenic lake characteristics, additional variables were included in the models. These included lake area, maximum lake depth, trophic state index (a measure of productivity), and the proportion of each lake watershed that contains agricultural, urban, or forested land. The data contain the following information (each row is an individual lake):

`DOW` unique 8-digit lake identifier

`yr` year of the survey

`numSc` number of docks scaled by shoreline perimeter in kilometers

`urban` proportion of land use in a lake's watershed as urban

`ag` proportion of land use in a lake's watershed as agricultural

`forest` proportion of land use in a lake's watershed as forest

`areaHec` surface area of lake in hectares

`depthM` maximum lake depth in meters

`tsi` trophic state index, 0-100 with higher values indicating more productive lakes

`allRich` total species richness

`submRich` submersed species richness

`sensRich` sensitive species richness

`emergFloatleafRich` emergent and floating-leaf species richness

The potential responses are the richness variables, which are counts of the number of different species present in the lake. The remaining variables apart from the lake identifier are potential predictors. For this problem, use `sensRich` as the response variable. Develop an appropriate (generalized) additive model, and summarize the results. Remember that the primary question is the role of human development on lake richness, and human development is mostly reflected by `urban`, `ag`, and `numSc`.

The data for this example were provided by Marcus Beck. I will email the data file to each of you. Please do NOT further distributed these data, or use them for any purpose other than for this course. If you save the file `sdata.rda` in the directory from which you start R, you can type:

```
load("sdata.rda")
str(sdata)
```