Assignment #3 — Due Friday, February 22, 2007, by 4:00 P.M.

Turn in homework in lecture, discussion, or your TA’s mailbox (just inside the main entrance to MSC). Please circle the discussion section you expect to attend to pick up this assignment.

311: Tues. 1:00–2:15  
312: Wed. 2:30–3:45  
313: Tue. 4:00–5:15

Instructions.— In a write-up of your solution to this assignment, you do not need to include graphs in your solution unless the questions specifies this explicitly. You do not need to show R code used to create graphs.

In questions that require you to fit a model and comment on estimated coefficients, your solution should describe the fitted model in an equation with words and units that is distinct from any R code used to fit the model. For example, from lecture notes, I could express the model for fitting the energy requirements for a bird of as a function of its mass from the model where I fit parallel lines as follows:

Energy (in Watts) = 6.02 + 0.0575 × Mass (in grams)

Your solution should then also include a summary of the R code you used to fit the model and summarize the results. You should edit the R output to eliminate material that is immaterial to the question. For example, if you use the summary function, you might choose to include in your solution the table of estimated coefficients, but you should eliminate the excess output such as quantiles of residuals.

Background.— A samara is the winged fruit from a tree that falls to the ground with a helicopter-like motion. A forest scientist is interested in predicting the terminal velocity of falling samaras based on disk loading, the mass of a samara divided by the area of the disk it passes through while spinning. A lower velocity could be related to a larger average dispersal distance allowing samara to drop further from the maternal tree, potentially decreasing competition for new resultant seedlings.

The scientist collected from 25–30 samara from each of seven species and measured the disk loading and terminal velocity of each. The species were white ash (Fraxinus americana), green ash (F. pennsylvanica), tulip tree (Liriodendron tulipifera), sugar maple (Acer saccharum), box-elder (A. negundo), red maple (A. rubrum), and silver maple (A. saccharinum). Samara from the two ash species and the tulip tree are bilaterally symmetric and roll as they spin. In contrast, the samara from the maples and the box elder (genus Acer) are asymmetric and they do not roll as they spin. Theoretical study of helicopter rotors suggests that the terminal velocity should be linearly related with the square root of the disk loading.

The data set for this assignment is on the course web page and is named samara.txt. The columns are:

1. species: the common name of the species of tree
2. genus: the scientific genus
3. velocity: the terminal velocity in meters per second
4. mass: the mass of the samara in milligrams
5. area: area of the disk while spinning in square-centimeters
6. loading: the ratio of mass to area
7. symmetry: a factor indicating if the samara are bilaterally symmetric or not
1. Examine the mass of the samara for each species both graphically and numerically. Summarize your observations. (Use R to solve, but no graphs or R code is necessary in your summary.)

2. Find the mean and standard deviation of the terminal velocity and the disk loading separately for each species. (Include R code for the calculation. Recall that the function `split()` will split a quantitative variable by the levels of a factor and that `sapply()` can be applied to the result of `split()` along with a function such as `mean()` or `sd()`.)

3. Plot terminal velocity versus the square root of disk loading. What features do you notice in the data? You should use a different color or plotting symbol for each species. Use the argument `groups=species` in `xyplot()`. (Bonus points if you add a key to your plot. See the bats.R example if you want to try this.) Include the plot with your solution.

4. Fit five regression models to predict velocity from the square root of disk loading and these additional explanatory variables:
   
   (a) none;
   
   (b) species;
   
   (c) species + species:sqrt(loading) interaction;
   
   (d) genus;
   
   (e) symmetry.

   For each model, write the equation that predicts the terminal velocity for a sugar maple samara in the style above. Find a numerical prediction for the terminal velocity of a sugar maple samara with disk loading value 3.0 mg/cm² for each model.

5. Use the regression model from (4e) above to predict the terminal velocity of an individual samara from each species where the disk loading is one standard deviation above the mean for that species. (Use the mean and sd found in the second question.)

6. Which model do you think best fits the data? How many parameters are needed to describe this model? Justify your response on the basis of your examination of the plotted data and estimated models.

Work to do, but not turn in.

- Read Chapters 5–6 of the textbook.