Midterm 2

Instructions:

1. The examination is due at 9:30 am at the beginning of lecture on **Thursday, May 1**.

2. For your solution for each problem, please use only one side of each piece of paper. You may write neatly or use a word processor. Begin the second solution on a new sheet of paper. Each separate solution should contain a **maximum of five pages, including computer output**. Many excellent solutions will use **fewer than the maximum allotment of pages**. You will likely create many graphs in your analysis. Only include graphs in your solution that are important to illustrate points. Do not exceed the page limit by appending many pages of graphs and computer output!

3. Begin each solution with a short paragraph that summarizes the main results and refers to the statistical evidence in support of your conclusions. This summary paragraph should be a well written, self-contained, description of your results that makes sense to a reader who has read the problem. This summary paragraph should not include computer output, tables, or too many details of the analysis.

   Follow the summary paragraph with a concise description of the data analysis and your thinking and reasoning to justify the analysis. Any computer output should be incorporated into your discussion. Only include computer output that is directly relevent to support your main conclusions.

4. You may use any books and notes that you desire, but the work that you turn in **must be completely your own**. In particular, **you may not discuss the examination with anyone other than the instructor**. Any questions you have can only be addressed to me.

5. Strive to provide a solution that addresses the main scientific questions of interest, and use the simplest and most straightforward statistical methods that are sufficient to this task. It is much more preferable to provide a simple sensible statistical analysis than to attempt to demonstrate the full range of your mastery of the most complex statistical methods.

6. The total exam score is 100 points. Each problem is equally weighted.

Problem 1:

Scientists in Costa Rica are interested in the process of forest regeneration in former plantations. A former plantation can be replanted with a fast-growing tree and then be left unmanaged for many years. The planted trees grow to form an overstory, and can be harvested. As time passes, new woody plants will grow underneath from seeds dispersed to the sites by natural means, such as animals, winds, and water. When the planted trees are harvested, the woody plants growing underneath can regenerate natural forest.

The scientists were interested in determining which of six commonly planted species of tree promoted the greatest forest regeneration as measured by the density of new woody plants in the understory. Data was collected at three sites, La Selva, Paniagua, and Quesada. The first was a former research plantation and the latter two were privately owned plantations. At each site, 30 plots were set up and five plots were planted for each species of tree. The six species were *Calophyllum brasiliense* (Cb), *Hieronyma alchorneoides* (Ha), *Terminalia amazonia* (Ta), *Virola koschnyi* (Vk), *Vochysia ferruginea* (Vf), and *Vochysia guatemalensis* (Vg). Nine years later, the plots were measured by counting the total number of woody plants in the understory a 4×4 square meter region within each plot.

The plots also differed by other characteristics. Trees were planted in regular arrays, but of different dimensions. For example, trees planted in a 2×4 array had a spacing of 8 square meters per tree while trees planted in a 4×8 array had a spacing of 32 square meters per tree. Most plots were on flat ground, but a few are sloped up to 10 percent. The drainage in most plots is good, but is poor in a few plots. The amount of light available to the understory was measured in each plot as a percentage of open canopy.

The data for this problem is in the file `regenerate.txt`. The variables are:

1. `stems`, the count of woody stems in a 16 square meter plot;
2. `site`, the name of the site;
3. `treatment`, the planted species of tree (abbreviated);
4. `spacing`, the number of square meters per planted tree;
5. `slope`, a factor with levels `flat` and `sloped`;
6. `drainage`, a factor with levels `good` and `poor`; and
7. `canopy`, the percent open canopy.

(a) The primary question of interest is to compare the regeneration measures of the six different species of planted trees. Prepare graphs that compare the regeneration measure for the treatments, indicating the sites as well. Do the different treatments behave similarly at each site?

(b) Create a model to predict the woody stem counts in plots using any important inputs. Justify your selection of the model.

(c) Use your model to create a confidence interval for the difference in regeneration measures between *Calophyllum brasiliense* (Cb) and *Terminalia amazonia* (Ta) at each site. Explain how you find these intervals.

(d) The site of a new former plantation will be planted with one of the six species with the hope of maximizing the forest regeneration. The new site is flat and has good drainage. Which tree would you recommend planting, and at what spacing? You will need to make some additional assumptions to answer the question. (For example, the canopy openness in the future will depend on the growth of the trees planted.) Briefly justify your recommendation.

Problem 2:

A researcher is interested in seeing how different light conditions (Long = 16 hr days, Short = 8 hr days) affects the concentration of an enzyme in golden hamsters in heart tissue. Ten hamsters are randomly placed into each of the two light treatment groups, for a total of twenty hamsters. After a treatment period of several weeks, the hamsters are sacrificed and enzyme concentration (mg/ml) is measured in four subsamples of heart tissue for each hamster. The data is in the file `hamster.txt`. The primary research question is to compare the effects of light on the enzyme concentration in heart tissue.

(a) Display a graph that highlights the primary research question.

(b) Fit and interpret a model that addresses the primary research question. Summarize your results.

(c) Find prediction intervals for the difference in concentrations of two heart samples from the same hamster in the long day treatment group and from the same hamster in the short day treatment group. Show how you arrived at the intervals.

(d) Find a prediction interval for the difference in mean concentrations of the four heart samples from two different hamsters, one in the long day and one in the short day treatment group. Show how you arrived at the interval.