**CPaT Stats Midterm - Week 5**

**Hard Copy – due Tuesday, Week 6, in Lab (hand in your Week 5 lab at the same time).**

There are four questions on this midterm (plus one optional). Questions 1 and 2 are “open-ended”. That means you could spend many hours working on them, but you have limited time, so I’ve limited the number of pages for each question. If you have been keeping up with labs, you should be able to do this in less than 6 hours.

This Take Home Exam should be an individual effort, but you are free to ask questions in class and lab, or email judyc if you get stuck. If there are clarifying questions that need to be answered I will email the whole class. Questions 3 and 4 are probably easier (and will be quicker than 1 and 2), so you might want to warm up on those.

**Question 1 (35%):** Designing a STELLA Model. On the fileshare, in /Handouts/Stats/MidTerm you will find a sub-directory (h2olos). It contains a description and diagram (h2olosModelSpecs) that describe a climate impacts model. The subdirectory also has additional background information that you may (but need not) refer to. Dominique has expressed interest in having a simplified STELLA model to illustrate the main points of her model. You have been tasked with working on that project! Your first steps as a modeler for Dominique are outlined below:

1. Read the description and study the diagram.
2. Do your best to produce a first cut STELLA (stock & flow) model of Dominique’s model.
3. Use STELLA to diagram your representation.
4. Print the STELLA model, as well as the Equations that it would generate.
5. Write down the 3-5 most critical questions you have for Dominique regarding the model
6. Write down the data that you would need from Dominique in order run your model.

As your answer to this question, turn in: 4,5,6; two pages max!

**Question 2 (35%):** Summary Statistics. On the fileshare, in /Handouts/Stats/MidTerm you will find not only this document, but also of the full 1kcs data set 1kcsStemMapAllforMidterm.xlsx. This is a very large study (almost 6000 rows)! You have just been hired as the data analyst on a large NSF project, and your first assignment is to prepare a 2-page summary of this data set.

You don’t have room (or time) to analyze every variable, but you might think about the most important variables defining this data set: Site ID, Species, DBH (diameter at breast height), tree height, live-status, crown base (the height from ground at which the tree’s branches have leaves).

My advice is to keep this simple, especially if you are short on time! It’s better to do a good job with a smaller data set than a sloppy one with a large data set. Pick a few variables to explore, or even just a few sites (that will make your data set smaller). Think about what might interest your audience. You might report summary statistics by Site, or by Species, or overall.

For the adventuresome among you, the crown data provides interesting opportunities. Crown radii essentially give you the projection of the tree crown onto the forest floor at North, East, South, West – if you were to draw a circle (or diamond) around the tree using these numbers, you would have the crown area; you could also calculate the crown volume by figuring the difference between tree height and crown base, and using the crown area).

Below I suggest a strategy for approaching this problem:

1. Take a little time (say 15 minutes) to peruse the data set and the 1kcs web site: <http://acdrupal.evergreen.edu/studycenter/1kcs> and get a feeling for the study. In particular, there you can (if you choose) download data for just one site.
2. Think about what you are interested in presenting, and decide which sites and which variables you want to work with. Write those down, along with your reason for choosing them. It’s OK to say – summary statistics for these variables (e.g., DBH and Height) looked easier to me!
3. Decide which summary statistics you want to run, and do that in JMP.
4. Prepare your two page report (see below).

Turn in your two page report on summary statistics for the 1kcs. Your report should say why you chose the sites and variables you chose, the summary statistics you ran (graphs, tables, etc.), and any preliminary ideas you have for further analysis.

**Question 3 (15%).** T-test. Plant scientists have developed varieties of corn that have increased amounts of the essential amino acid lysine. In a test of the protein quality of this corn, an experimental group of 20 1-day old chicks was fed a ration containing the new corn. A control group of another 20 chicks received a ration that was identical except that it contained normal corn. The data set contains the weight gains (in grams) after 21 days. The data for this is in the excel file Corn-LysineStudy.xlsx. (G. L. Cromwell et al., “A comparison of the nutritive value of opaque-w, floury-w and normal corn for the chick,” *Poultry Science*, 47 (1968), pp 840-847.)

1. Present the data graphically. Are there outliers or strong skewness that might prevent the use of the t-test?
2. State the null and alternative hypotheses for a statistical test of the claim that chicks fed high-lysine corn gain weight faster. Carry out the test, and report the result.
3. (optional) give a 95% confidence interval for the mean extra weight gain in chicks fed high-lysine corn.

Your answer to this question should be no more than 1 page!

**Question 4 (15%).** 1-way ANOVA. The presence of harmful insects in farm fields is detected by erecting boards covered with a sticky material and then examining the insects trapped on the board. To investigate which colors are most attractive to cereal leaf beetles, researchers placed six boards of each of four colors in a field of oats in July. The data in ColorsAndBeetles.xlsx shows the number of beetles trapped in each of the 24 boards. (from M. C. Wilson and R. E. Shade, “Relative attractiveness of various luminescent colors to the cereal leaf beetle and the meadow spittlebug,” *Journal of Economic Entomology*, 60 (1967), pp 578-580.)

1. Make a table of means and standard deviations for the four colors and present the data grapically.
2. State the null and alternative hypotheses for an ANOVA on these data, and explain in words what the ANOVA will test in this setting.
3. Run the ANOVA (in JMP or resampling) and report your results.

Your answer to this question should be no more than 1 page!

**Question 5.** Optional. For the 1kcs study, suggest 1 to 3 interesting ANOVA tests that would elucidate the differences among the different sites or species. Say why you think those tests would be interesting, specify the variables you would use (dependent and independent), your research hypothesis, and your null and alternative statistical hypotheses. If you still have time to work on this exam, try running those ANOVAs and report your results (or say what difficulty you ran into. (2 pages max)