Final—Due Wed 17 May at noon in my mailbox

The final consists of several problems. Computer printout is attached to save you from fighting for computer access. In fact, I would rather you NOT attempt more detailed computations, but instead THINK about what you have. If relatively simple calculations are called for (addition, multiplication), I expect you to do them by hand. If more involved calculations would be necessary to "answer" the question, then your task is to explicitly identify what is needed and why. (Note that if you have an parameter estimate and an estimate of its standard error, I expect you to be able to conduct inference. However, if the estimate or its standard error are computed incorrectly, show why, and what would be correct.)

Prepare separate reports for each problem geared toward a statistician with the background of Stat 850. Use a combination of plots (diagnostic, interaction, other?) and small summary tables to supplement each report. Label plots and tables clearly. Please keep it as short as possible – the total length should be shorter than 15 pages (longer finals may receive less attention...). Typing is not required, but neatness and readability is!

1. A plant scientist is interested in finding out the location of genes which control various tomato plant attributes. He is using molecular markers to do this. He thinks he has located a major gene for "brix" (`bx`) close to marker `tg99`.

The original experiment covered two years (`yr`), with 93 unique plant entries (`entry`). There were several replications each year. Unfortunately, the data now available to the scientist consists of the mean brix (`mbx`), averaged across the replicates for each year, with some missing. [The raw data are in notebooks halfway around the world!] The scientist believes a square transformation (`mbx2`) is reasonable, and has presented that data in that way.

Nevertheless there are still 2 years of data for most entries. The marker `tg99` can be used to classify entries into one of three categories, 1 = parent A, 3 = parent B, 2 = hybrid of A and B (and . = missing marker value). The scientist is particularly interested in the "additive effect" (parent A – parent B) and the "dominance effect (hybrid – mean of parents). Note that if the dominance is zero, then the hybrid would be halfway between the two parents.

Your task is to compare the three marker "genotypes" over the two years: are they different? is the pattern consistent over years?

2. An experiment on potato growth was conducted in the biotron, a national facility for controlled experimental environments located on the Madison campus. The researcher wanted to compare 4 varieties of potatoes with 5 replicates. Unfortunately he could only fit 4 plants (one of each variety) in a growth chamber at a time. Further, only one growth chamber was available. Thus he decided to replicate the experiment by running 4 plants at a time on 5 separate occassions (20 plants total) over the Spring semester.
a) describe the treatment structure and the design structure.
b) write down the model with all assumptions and side conditions.
c) Analyze the data. Include in your analysis the ANOVA table, E(MS), hypotheses and test statistics, and any followup analysis (comparing means?) or interpretation (problems, cautions?) as appropriate.

3. A dairy scientist wanted to compare milk yield for cows fed alfalfa (`trt=A`) or mixed grains (`trt=M`). The cows were placed in paddocks (`pad`), grazing areas where they could feed freely. Based on previous studies, it is usually important to distinguish between milk production for cows having their first offspring (`parity=H`) from those that are more experienced (`parity=C`). You may assume that cows were randomly assigned to paddocks, and that the milk production for one cow is not dependent on behavior of other cows in the same paddock.

Milk production was measured over 13 weeks. You have available the results of a repeated measures analysis using `proc glm`. This includes the various types of analyses we discussed in class. Write down your model (or models) and accompanying assumptions. It may help to first consider a single week's measurement, and build up to the repeated measures by way of split plot. Imagine at each step that you have only one measurement per cow. Interpret your results, indicating how the various approaches complement each other. You may find that you have to do some hand calculations to augment the "automatic" output, but these should be minor.
If you need access to the data and code, they can be found in `fin?.dat`, `fin?.sas` and `fin?.prt` in the usual place.