Stat 402 A - Spring 2008 - Final exam, (Takehome)

Due: Friday, May 9, by 5 pm. There are three questions, each with multiple parts.

Reminders: **You are not allowed to ask classmates or friends for help**. Please ask me, not a classmate, for help with SAS, even if you are working next to that classmate in the same computer room.

You are allowed to use your book(s) and your notes. You are always welcome to e-mail me with questions or requests for clarification. I will not provide SAS code, but I am very happy to answer questions about syntax, output, or errors. If you do not understand the context for each problem, ask for clarification.

Some questions should be easy, some are harder, and a few require you to integrate ideas, which many folks find hard. You don't have to get a perfect score to do well on the exam. I adjust my grading scheme for the difficulty of the exams.

Your answers may be written or typed. You may submit just your answers. However, it helps me to see your SAS code and output. That way I can give partial or nealy complete credit if you're thinking correctly but not asking SAS for the right thing.

I grade anonymously. Please **do not put your name on the front** of your answers. **Instead, put your name on the back of the last page** of the written/typed answers and make sure all pages are securely attached.

Arrange your answers in the following order:

Answer to question 1, answer to question 2, answer to question 3,
your name on the back of the last page of question 3,
SAS for Q 1, SAS for Q 2, SAS for Q 3. your name again on the back of the last page of SAS output

1. 30 pts. The data in apple1.txt and apple2.txt are from an experiment on apple orchards. The treatments are 6 different types of ground cover around the apple trees. Five treatments (A-E) are new experimental ground covers; treatment O is the ground cover in current use. The investigators want to know if any of the experimental ground cover treatments will increase apple yield.

   The experimental orchard was divided into four blocks. Treatments were randomly assigned to groups of 4 trees. The response, POST, is the total weight of apples (in pounds) produced by that group of trees after the treatments were established. You also have available the total weight of apples produced the year before the treatments (PRE).

   The data in apple1.txt has one line of data per e.u. The data in apple2.txt has the PRE information and POST information on two separate lines.

   (a) 10 pts. The investigators decide to compare treatment means for apples trees producing the same PRE-treatment yield. Write out the skeleton ANOVA table for the appropriate analysis.

   (b) 5 pts. Is there any evidence that the ground cover treatment influences yield, when treatments are compared at the same PRE-treatment value? Report the appropriate test statistic and p-value.

   (c) 5 pts. Estimate the predicted mean yields after groundcover treatments for trees producing 350 pounds before the treatments were applied.

   (d) 5 pts. Which treatments are effective, i.e. produce significantly more apples than the current treatment (treatment O)? Include test statistic(s) and p-value(s) in your answer. You do not need to worry about multiple testing adjustments.

   (e) 5 pts. After doing the above analyses, the investigators decide to fit a heterogeneous regression lines model: $\text{POST}_{ij} = \beta_{0i} + \beta_{1i}\text{PRE}_{ij} + \epsilon_{ij}$, where i indicates the treatment and j indicates an observation. In this model, the relationship between PRE and POST is described by straight lines with different slopes and different intercepts for each treatment. Use this model to test whether the treatments really do have different slopes. In other words, test H0: all treatments have the same slope.

2. 40 pts. Insulin is an important regulator of animal metabolism. The following data and problem are based on a study of the effects of diet on blood plasma insulin level in steers. Three different diets (DIET A: 2.4 Mcal/kg, DIET B: 2.7 Mcal/kg, and DIET C: 3.0 Mcal/kg) and nine steers were used in the experiment. Because the investigators expected a large variation between steers, the experiment was designed as a 3 x 9 (periods x steers) latin rectangle, where each diet was fed to 3 steers in period 1, 3 different steers in period 2, and 3 steers in period 3. Over the three periods, each steer was fed each of the three diets. Sufficient washout time was provided between experimental periods so there is no concern about carryover from one treatment to another. Each period was 21 days long. Within each period, blood plasma insulin concentration was measured 7 times on each cow, at days 3, 7, 10, 14, 17, and 21. Although baseline data (day 0) were also collected for each cow, that data is ignored in this problem. The data are in insulin.txt on the class web site.

   The investigators are interested in knowing:
   Whether diet had any effect on the mean insulin concentration over the 21 day period.
   Whether the effects of diet were the same throughout the 21 day period. The investigators expected the effect of diet to increase throughout the experiment. On what days was there some effect of diet?

   (a) 10 pts. Write out the skeleton ANOVA table for a split-plot in time approach to analyze these data. Indicate which terms in the ANOVA table are fixed and which are random.

   (b) 10 pts. Is the split-plot in time approach reasonable for these data? If not, what correlation model is more appropriate? Briefly explain your choice. You only need to consider the correlation models we've discussed in class.

(c) 10 pts. Answer the investigators two questions:
Did diet have any effect on the mean insulin concentration over the 21 day period.
Were the effects of diet the same throughout the 21 day period?
Include in your answers the appropriate test statistic(s) and p-value(s), appropriate table(s) of means, and a short summary of the pattern in insulin concentration.

(d) 5 pts. The investigators expected the effect of diet to increase throughout the experiment. This question is similar to, but more focused than, their second question (were effects of diet the same at all times?). The focused question suggests using the slope (Y = insulin concentration, X = day) as a summary statistic. Test whether the diets have the same mean slopes. Report your F statistic and p-value.

(e) 5 pts. The investigators decide to confirm what they see from the previous test by testing the effect of diets separately on each day. On which days was there a statistically significant effect of diet? Report test statistic(s) and p-value(s).

3. 30 pts. The data in stream.txt are made-up data based on the LINX II experiment. The context is Nitrogen transport from agricultural fields in the upper Midwest down the Mississippi River to the Gulf of Mexico, where excess N is having severe environmental effects. This experiment studied whether streams serve simply as pipes transporting Nitrogen downstream, or whether biological processes in streams retain some (perhaps all) of the Nitrogen instead of moving it downstream. The LINX experiment is a multi-location comparison of three types of streams (undisturbed, agricultural, and urban). At one location, researchers studied 9 streams; 3 were undisturbed, 3 were agricultural, and 3 were urban. This basic experiment was repeated at 8 different locations scattered around the US and Puerto Rico. There are a total of 72 observations (9 streams at each of 8 locations). The response is Ammonium uptake distance; this was measured on each stream. There are a lot of details of measurement and methodology that are irrelevant for this problem. For interpretation, it may help to know that a large uptake distance means that the stream is acting like a pipe and just moving N downstream. A short uptake distance means that a stream is retaining N.

(a) 5 pts. The investigators are interested in comparing the three types of streams (treatments), while accounting for possible differences between the eight locations and the possible inconsistency of treatment effects across the 8 locations. Write out an appropriate skeleton ANOVA table for this study, showing sources of variation and d.f.

(b) 5 pts. If you are interested in broad sense inference about the differences between streams, which terms in the ANOVA table should be considered fixed and which should be considered random?

(c) 5 pts. The investigators would like to estimate the average difference between undisturbed and agricultural streams at these eight specific locations. Estimate the mean difference and the appropriate standard error.

(d) 5 pts. Test whether the differences between stream types are large relative to the consistency of treatments across locations. Report the appropriate F statistic and p-value.

(e) 5 pts. When you examine residuals, you realize that the data should be log transformed prior to analysis. Repeat the test in part 3d after log transforming the responses. Are the conclusions the same? Explain why or why not.
Hint: it may help to draw an interaction plot showing means for each stream type and location as a function of average uptake distance for each location. Do the same after log transformation.

(f) 5 pts. The investigators are interested in inferences to streams throughout the US. Estimate the average difference in log-transformed responses between undisturbed and agricultural streams. Report the estimate and the appropriate s.e.

That's all! I've enjoyed working with you this semester. Thanks for your effort and thoughtful questions. Please sign the appropriate statement below and include the page after your answers.

Go celebrate and have a wonderful summer!

_____ I completed this final without assistance from friends or classmates

_____ I received the following assistance from friends or classmates (please describe on a separate sheet)

Signed: _____