Stat 402B   Final Exam   May 4, 2001

1. A gardener is interested in studying the relationship between fertilizer and tomato yield. The gardener has two gardens (1 and 2). He divides each into 9 plots. Three fertilizer application rates (3, 5, and 7 units/acre) are assigned to the plots in garden 1 in a completely randomized fashion. The same three fertilizer application rates (3, 5, and 7 units/acre) are assigned to the plots in garden 2 in a completely randomized fashion. Thus there are three plots for each combination of garden and fertilizer application rate. After some initial analyses, the gardener decides to base his analysis on the following SAS code and output.

```
proc glm;
  class garden;
  model yield=garden rate garden*rate / solution;
run;
```

```
The GLM Procedure
Dependent Variable: yield
                                  Sum of
Source          DF           Squares     Mean Square    F Value    Pr > F
Model            3       58.88888889     19.62962963      27.33    <.0001
Error           14       10.05555556      0.71825397
Corrected Total 17       68.94444444

Source          DF       Type I SS       Mean Square    F Value    Pr > F
garden           1        2.72222222      2.72222222       3.79    0.0719
rate             1       52.08333333     52.08333333      72.51    <.0001
rate*garden      1        4.08333333      4.08333333       5.69    0.0318

                                    Standard
Parameter          Estimate           Error      t Value      Pr > |t|
Intercept          -1.11 B       0.90993803        -1.22        0.2422
garden       1      3.69 B       1.28684670         2.87        0.0123
garden       2      0.00 B        .                  .           .
rate                1.33 B       0.17299494         7.71        <.0001
rate*garden 1      -0.58 B       0.24465179        -2.38        0.0318
rate*garden 2       0.00 B        .                  .           .
```

(a) Note that *rate* was not included in the class statement. What would the *Model* and *Error* DF change to if *rate* were included in the class statement?

(b) Estimate the equation of the regression line relating yield to fertilizer application rate in garden 1.

(c) Estimate the equation of the regression line relating yield to fertilizer application rate in garden 2.

(d) Is there a significant difference between the slopes of the two regression lines? Give an appropriate test statistic, p-value, and conclusion at the 0.05 level.

(e) Suppose the gardener were to apply 7 units of fertilizer per acre to all plots in both gardens. Which garden would have the higher expected yield?

2. An experiment was conducted to compare time to germination of 3 varieties of carrots grown in 2 types of potting soil. Each variety was planted in 4 pots of each potting soil so that a total of 24 pots were used for the experiment. The 24 pots were randomly positioned in a greenhouse. During the experiment, several pots were accidentally knocked off the greenhouse bench so that time to germination was only measured for 16 of the original 24 pots. The available data are provided in the table below. Following the table is some relevant SAS code and output.

	Variety		
Potting Soil	1	2	3
1	7,9	3,5,4,2	10
2	7,6,6,5	2	7,8,8,7

```
proc glm;
  class soil var;
  model y=soil var soil*var;
run;
```

```
                                         Sum of
Source              DF         Squares      Mean Square     F Value     Pr > F
Model                5     74.00000000      14.80000000       14.80     0.0002
Error               10     10.00000000       1.00000000
Corrected Total     15     84.00000000

Source              DF       Type I SS      Mean Square     F Value     Pr > F
soil                 1      1.01587302       1.01587302        1.02     0.3373
variety              2     72.58412698      36.29206349       36.29     <.0001
soil*variety         2      0.40000000       0.20000000        0.20     0.8219

Source              DF     Type III SS      Mean Square     F Value     Pr > F
soil                 1     11.07692308      11.07692308       11.08     0.0076
variety              2     62.14545455      31.07272727       31.07     <.0001
soil*variety         2      0.40000000       0.20000000        0.20     0.8219
```

(a) Based on the output above, is there a significant difference between soil types? Give an appropriate test statistic, p-value, and conclusion at the 0.05 level.

(b) Compute LSMEANS for each soil type.

(c) If someone asked you to recommend the potting soil that will lead to shortest germination time, which type of potting soil (1 or 2) would you recommend? Explain briefly.

(d) Let μ_{ij} denote the mean germination time for plants of variety j potted in soil i. Give a 95% confidence interval for

$$(\mu_{11} + \mu_{21})/2 - (\mu_{12} + \mu_{22})/2.$$

(e) What can you conclude from the confidence interval you constructed in part (d) with regard to germination time of carrot varieties?

3. A statistics professor has a rowing machine with five air resistance settings (1, 2, 3, 4, and 5). He rows for 30 minutes every weekday morning. The machine has a computer that calculates the distance traveled (in meters) during the 30-minute row. Similar to the gearing on a bicycle, the highest setting (5) provides the most resistance on each rowing stroke but also maximizes the distance traveled with each stroke. The lowest setting (1) provides the least resistance but minimizes the distance traveled with each stroke. The professor decides to conduct an experiment to determine the single setting that will allow him to row the farthest during his morning row. The professor plans to record the distance traveled on each 30-minute morning workout every weekday for the next 5 weeks. These 25 workouts will serve as the experimental units. The treatments consist of the 5 air resistance settings.

   (a) Name the experimental design that you would use for this experiment, and give an example of one possible random assignment of air resistance settings to morning workouts by entering treatment numbers in the table below.

	Week				
Day	1	2	3	4	5
Monday					
Tuesday					
Wednesday					
Thursday					
Friday					

   (b) Briefly explain why you chose this design.

4. A researcher would like to investigate the effect of feeding high-oil corn to one-year-old steers. Four feedlots, each with 5 pens, are available for use in the experiment. Each pen holds 6 steers. For practical reasons, all six steers in a single pen must be fed the same diet. Five diets are selected for use in the experiment. The diets are (A) a traditional corn diet, (B) a diet with 25% high-oil corn, (C) a diet with 50% high-oil corn, (D) a diet with 75% high-oil corn, and (E) a diet with 100% high-oil corn. The researcher is planning to measure the ribeye area of each steer at slaughter.

   (a) Name the experimental design that you would use for this experiment, and give an example of one possible random assignment of diets to pens by entering treatment letters in the table below.

	Pen				
Feed Lot	1	2	3	4	5
1					
2					
3					
4					

   (b) What are the experimental units in this experiment?

   (c) What are the observational units?

   (d) Provide the SOURCE and DF columns of the ANOVA table for the analysis of this experiment.

   (e) Write down an appropriate linear model for this experiment. Briefly explain the meaning of each term in the model. Be sure to explain any assumptions that you are making about random effects. (If you prefer, you may simply provide appropriate SAS *proc glm* or *proc mixed* code for the analysis of this data. Be sure to include *class*, *model*, and *random* statements.)

5. A team of researchers is studying erosion on highway construction sites. The banks of a freeway overpass can be quite steep. If it rains hard on a recently built bank, soil can wash down and contribute to pollution in nearby streams. The researchers would like to compare the effectiveness of four different grasses at preventing erosion. Ten freeway banks are available for study. Each bank is divided into four adjacent plots. The 4 grasses are randomly assigned to the plots in such a way that all 4 grasses appear in random order on each bank. The research group has a rainfall simulator that provides a fixed amount of rain simultaneously to all 4 plots on a given bank. Five of the ten banks are randomly selected to receive the artificial rainfall one month after planting. The other five banks receive the artificial rainfall three months after planting. The amount of soil washed off each plot after the artificial rain is the response variable.

   (a) Provide the SOURCE and DF columns of the ANOVA table for the analysis of this experiment.

   (b) Add a third column to the ANOVA table you started in part (a). Write "random" in each row corresponding to a random term. In the row for each fixed term, write the name of the error term that should be used to test the significance of the fixed term.

   (c) An appropriate model was fit to the data using SAS. A significant *time*-by-*grass* interaction was discovered. (Here *time* denotes the factor with levels *one month after planting* and *three months after planting*. A portion of the output regarding a test of sliced effects is provided below.

```
Tests of Effect Slices

Effect        time        F Value    Pr > F
time*grass    1 month        1.51    0.2368
time*grass    3 months      24.60    <.0001
```

      i. What are the numerator and denominator degrees of freedom corresponding to the F statistics presented in the output?

      ii. What can you conclude from these F tests?