**STATISTICS QUALIFYING EXAM - TAKE HOME**
**Due Friday, January 14, 2000. Return to Math and Stat dept office**

*Directions:* The answer to each problem should be word-processed, double spaced. An appendix is allowed for each problem but will be examined only at the discretion of the graders. The better constructed your appendix, the more likely it is to get examined.

You may not consult any other person when working on this exam or discuss your exam with anyone else regardless of whether or not the person is taking the exam.. You may use only your class notes for the exam. You may use any available books or web resources for the exam.

1. The data below consist of survival times for a number of patients with a certain type of cancer along with the values of five potentially important prognostic variables (covariates, explanatory variables). The variables are:

   - $y$: survival time in days
   - $x_1$: log blood urea nitrogen
   - $x_2$: platelets, 0=abnormal, 1=normal
   - $x_3$: percent lymphocytes
   - $x_4$: protein in urine at diagnosis: 1= present, 2=none
   - $x_5$: serum calcium

   Data on the first 2 patients is: The full data set can be found at *LIST IT*

| Patient | $y$ | $x_1$ | $x_2$ | $x_3$ | $x_4$ | $x_5$ |
|---|---|---|---|---|---|---|
| 1 | 1.25 | 2.2175 | 1 | 0 | 2 | 10 |
| 2 | 1.25 | 1.9395 | 1 | 0 | 1 | 18 |

   Analyze these data giving the appropriate model, choosing the important explanatory variables. Be sure to examine residuals, make transformations if necessary, and develop a model to predict the survival time of such patients.

2. Precast concrete paving slabs, known as slabs, are the usual surface for pavements. These have been in use since the 1890s when they started to replace the high cost stone flags. They were popular because they were cheaper to buy and to lay. The savings in construction costs were due mainly to the higher dimensional accuracy of the concrete flags. Unfortunately, pavements are sometimes over-ridden by heavy vehicles and this can damage the surface. The costs of pavement maintenance is considerable and there has been a lot of research into methods for limiting the damage caused by misuse. One such experiment is to analyze the effect of flag size and joint width on the load transfer of pavements. The data are given below:

   (a) Analyze these data. Give an appropriate analysis of variance table.

   (b) Identify the model clearly explaining all of the variables.

| Flag Type | Joint 3 mm | Joint 6 mm | Joint 10 mm |
|---|---|---|---|
| A | 0.200 | 0.120 | 0.121 |
| B | 0.530 | 0.280 | 0.180 |
| C | 0.327 | 0.142 | 0.158 |

Table 1: Maximum load transfer (tons) of flags.

(c) What can you say about flag types and joint widths vs. load transfer?

(d) Suggest any way to improve this experiment. Be precise in your answer and write down exactly how you would like to see it redone.

(e) If you were asked to analyze these data using regression analysis, how would you do it? (Just explain, don't do it!) In this case, what would the coefficient of $B$ estimate?

(f) In general, what is the difference between a RCB design and a CRD design with two factors?

(g) Why randomize? List two reasons and explain.

3. Suppose that a $2^3$ experiment is going to be conducted in 2 runs. The experimenter decides to arrange the design in 4 blocks of 2 runs so as to do as little damage as possible to the estimates of important effects. Suppose that he decides to confound the contrasts $ABC$ and $BC$.

(a) Write out the experiment arranged in 4 blocks.

(b) Are there any serious problems with this design? Justify your answer in detail.

(c) If you believe there are problems with this design, suggest a better scheme and write out the arrangements for your scheme.

(d) Analyze the following data from a $2^3$ factorial with 2 replicates. This data in on a study of the effect of percent impurity (A), operating pressure (B), and line speed (C) on the fill volume of a mixer for a composite material used in road construction. Two levels of impurity are used. Give the ANOVA table as part of your analysis. The data for fill volume are:

| | Operating pressure (B) | | | |
|---|---|---|---|---|
| | 25 psi | | 30 psi | |
| Percent | line speed (C) | | line speed (C) | |
| Impurity (A) | 200 | 250 | 200 | 250 |
| 10 | -3, -1 | -1, 0 | -1, 0 | 1, 1 |
| 12 | 0, 1 | 2, 1 | 2, 3 | 6, 5 |