METHODS QUALIFYING EXAM

JANUARY 2003

INSTRUCTIONS:

1. DO NOT put your NAME on the exam. Place the NUMBER assigned to you on the UPPER LEFT HAND CORNER of EACH PAGE of your exam.
2. Please start your answer to EACH QUESTION on a SEPARATE sheet of paper.
3. Answer all the questions.
4. Be sure to attempt all parts of every question. It may be possible to answer a later part of a question without having solved the earlier parts.
5. Be sure to hand in all of your exam. No additional material will be accepted once the exam has ended and you have left the exam room.

QUESTION 1

A study was conducted to compare different methods of reducing the concentration of cedar flies (insects that inhabit cedar trees). The reduction methods consisted of two types of chaining (cutting the trees down) along with burning all the trees, grass and leaf litter. (Cedar flies overwinter in the leaf litter.) The study design had the following "factors":

1. There were four different geographical regions containing cedar trees.
2. Within each region, there were two techniques for cutting the trees down (smooth chaining and elevated chaining) plus a control (no chaining) assigned at random to one of three groves. The chaining was done prior to collecting the data in the first year.
3. The study was conducted over a five-year period. The treatment assigned to a grove was the same every year. A sample of cedar flies was taken at fixed periods during the summer months and the average number of flies was recorded for each treatment (that is, one assessment of fly concentration for each grove for each year).
4. The time frame for years was:
   - Year 1: In the spring the treatments (smooth chaining, elevated chaining, and no chaining) were applied to the appropriate groves. Data were collected during the summer.
   - Year 2: No additional treatment of the groves. Data was collected during the summer.
   - Year 3: No additional treatment of the groves. Data was collected during the summer.
   - Year 4: For the sites with chaining, the litter was burned in the spring. Nothing was done to the control sites. Data was collected during the summer.
   - Year 5: No additional treatment of the groves. Data was collected during the summer.

There are a total of 60 data values.

(A) Give an appropriate ANOVA table for these data, showing sources of variation, degrees of freedom, and expected mean squares.

(B) What hypotheses would you test? Show the appropriate F-ratios for each hypothesis using the names of the mean squares.

(C) Are there any other tests and/or estimates you would conduct in addition to these F-tests? If so, describe them.

QUESTION 2

In both analysis of variance and regression analyses when the sample sizes are relatively small, the model conditions become crucial in obtaining valid inferences. In most situations, the residuals are used to assess whether these conditions hold for a given data set. Describe methods for using the residuals to assess the validity of each of the following conditions.

(A) Normality

(B) Independence

(C) Discuss three different procedures for testing the equality of variance. Give the advantages and disadvantages of each procedure.

QUESTION 3

Suppose 250 independent pairs, $(X_1, Y_1), \cdots, (X_{250}, Y_{250})$, are randomly selected from a population having an underlying quadratic relationship, $Y_i = \beta_o + \beta_1 X_i + \beta_2 X_i^2 + \epsilon_i$, where ϵ_i's are i.i.d. random variables with mean 0 and variance σ^2. The researcher suspects a nonlinear relationship between X and Y and decides to fit a cubic polynomial regression model to the data. Let $\mathbf{X_{[3]}}$ and $\mathbf{H_{[3]}}$ denote the design matrix and the "hat matrix" for this cubic polynomial regression. Also, let $\mathbf{1}$, $\underline{\mathbf{x}}$, $\underline{\mathbf{x^2}}$, and $\underline{\mathbf{x^3}}$ denote the columns in $\mathbf{X_{[3]}}$.

(A) Show that $\left(\mathbf{X^t_{[3]}X_{[3]}}\right)^{\mathbf{-1}}\mathbf{X_{[3]}1} = (\mathbf{1}, \mathbf{0}, \mathbf{0}, \mathbf{0})^{\mathbf{t}}$ and that $\left(\mathbf{X^t_{[3]}X_{[3]}}\right)^{\mathbf{-1}}\mathbf{X_{[3]}}\underline{\mathbf{x^2}} = (\mathbf{0}, \mathbf{0}, \mathbf{1}, \mathbf{0})^{\mathbf{t}}$

(B) Suppose one goal of the study is to estimate the mean of Y when $X = x_o$ for a given x_o. Is it possible to obtain an unbiased estimate of the conditional mean even though the wrong model has been fitted to the data? Use (A) to explain your answer.

(C) Suppose one also wants to provide a prediction interval for a possible new observation Y^* when $X = x_o$ using the fitted cubic model. Would the mis-specification of the model play a role in this prediction? Are there any other issues that one should consider before constructing such an interval?

QUESTION 4

Three different alloys were prepared with four separate castings for each alloy. Two bars from each casting were tested for tensile strength. The data are tensile strengths of the individual bars. The data is given below:

	Castings				
Alloys	1	2	3	4	Mean
A	13.2	15.2	14.8	14.6	14.7
	15.5	15.0	14.2	15.1	
B	17.1	16.5	16.1	17.4	16.7
	16.7	17.3	15.4	16.8	
C	14.1	13.2	14.5	13.8	14.1
	14.8	13.9	14.7	13.5	

(A) Write a linear model for the experiment. Make sure to identify which terms are fixed and which terms are random.

(B) Complete the following ANOVA table:

Source	DF	SS	MS	Expected Mean Squared
Alloys		29.38		
Castings		4.68		
Bars		4.32		

(C) Is there a significant difference in the mean tensile strength of the three alloys?

(D) Compute 95% confidence intervals for the mean tensile strength of the three alloys.

(E) Proportionally allocate the variance in tensile strengths to its components.