Methods Qualifying Exam

August 2000

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.

PROBLEM #1:

Weight gain in the first 3 months after birth is important for new born infants. A pediatrician wishes to test a new feeding formula to determine if it will cause greater weight gain in new born infants than the standard formula.
From her records she finds that the first 3 months weight gains of single birth infants on the standard formula have the following characteristics:

$$\mu_S = 15oz. \quad \text{and} \quad \sigma_S = 6oz.$$

On the other hand, the first 3 months weight gains of identical twins on the standard formula have the characteristics:

$$\mu_T = 12oz., \quad \sigma_T = 6oz. \quad \text{variation between sets of twins, and}$$

$$\rho = .8 \ \text{(correlation in weight gain of identical twins)}$$

She wants to run a 3 month experiment on a group of infants to test the new formula versus the old formula. She has decided to use a 5% probability of Type I error, and wishes to be able to detect a 3 oz. increase in weight gain with 90% probability.

(a.) What sample size must she use if she does the experiment with a CRD of single birth infants? State any assumptions you are making in this calculation.

(b.) What sample size must she use if she does the experiment with a RCBD with pairs of identical infants? State any assumptions you are making in this calculation.

(c.) Discuss the relative merits of the two experiments in terms of practicality and of her basic goal.

PROBLEM #2:

I. Consider the model

$$\text{Model} \quad (1): Y_{ij} = \beta_0 + \beta_1 X_i + \epsilon_{ij}$$

for $i = 1, \cdots, n$ and $j = 1, 2$ where we have 2 replicates at each vector $(1, X_i)$ of independent variables. Assume that ϵ_{ij} are uncorrelated random variables with mean zero and variance σ^2. Consider also the model

$$\text{Model} \quad (2): \bar{Y}_{i.} = \beta_0 + X_i\beta_1 + \bar{\epsilon}_{i.}$$

where $\bar{Y}_{i.}$ is the mean of the 2 replicates and $\bar{\epsilon}_{i.}$ is the corresponding error.

(a.) Find the ordinary least squares estimator for (β_0, β_1) in Model (1) and compare it with the weighted least squares estimator of (β_0, β_1) in Model (2).

(b.) Find a transformation in Model (2) [rescaled model from model (2)] such that the ordinary least squares estimates of (β_0, β_1) in this rescaled model coincide with the estimates you got from part (a.).

II. Let θ $(0 \leq \theta \leq 1)$ denote the probability of pain at dose x of a certain drug. One model for θ is to take

$$\theta = \int_{-\infty}^{x} f(u)du$$

where $f(u)$ is a probability density function. If $f(u)$ represents the extreme value distribution

$$f(u) = \beta \exp[(\alpha + \beta u) - \exp(\alpha + \beta u)], \qquad -\infty < u < \infty.$$

(a.) Find $\log\{-\log(1 - \theta)\}$.

(b.) Let m_i, $i = 1, 2, \cdots, n$ denote the number of patients exposed at dose level x_i and y_i denote the number of patients relieved from pain at that dose level. Show that for the logistic link

$$\theta = \frac{e^{\alpha+\beta x}}{(1 + e^{\alpha+\beta x})}$$

the maximum likelihood equations for estimating α and β can be written as

$$\sum_{i=1}^{n} (y_i - m_i\hat{\theta}_i) = 0,$$

$$\sum_{i=1}^{n} (y_i - m_i\hat{\theta}_i)x_i = 0$$

where

$$\hat{\theta}_i = \frac{e^{\hat{\alpha}+\hat{\beta}x}}{(1 + e^{\hat{\alpha}+\hat{\beta}x})}.$$

PROBLEM #3:

I. There are many private chemical laboratories in Texas that perform analyses of food products. One of the tests is the determination of the amount of cadmium in fish oil. Each laboratory has its own procedure for the assay, and the state regulatory agency wants to determine if a standard method would be more accurate.

> To determine the adequacy of the lab procedures, the agency took a homogeneous supply of the fish oil (ten gallons), divided it into 2 five gallon portions, and enhanced one of the five gallon portions with additional cadmium. Fifteen laboratories were randomly selected from a list of hundreds of qualified laboratories, and each was sent a sample of the enhanced and unenhanced oils. Each lab was to split each sample into two portions and analyze the portions using their own method and the new, standard method. Therefore, each lab was to perform four analses.

(a.) Write the model for the data of this experiment.

(b.) Determine the expected mean squares for the components of the AOV.

(c.) The interaction between the lab and the analysis method could be important in the detection of whether the old methods that the labs were using were giving equivalent results. Can this design detect such an interaction? If not, how would you modify the design so that it could detect the interaction?

II. After conducting the F-tests for main effects and interactions in a CR factorial experiment involving two factors A and B, the experimenter wanted to further investigate the experimental data. For each of the following questions given on the next page, select a **ONE** letter from the list at the bottom of the page which is the best solution to each of the following situations. **Place your selection** in the space to the **left** of each situation.

SITUATION:

.... 1. The A*B interaction is significant. The experimenter wants to compare the levels of factor A, which are unequally spaced numerical values.

.... 2. The A*B interaction is not significant. The researcher selects several contrasts in the levels of a qualitative factor A based on the information observed in the experimental data and then wants to test whether the contrasts are significant.

.... 3. The A*B interaction is not significant. The levels of both A and B are fixed qualitative levels. The experimenter wants to compare the levels of factor A.

.... 4. The A*B interaction is significant. The levels of A are randomly selected, levels of B are fixed qualitative levels. The experimenter wants to perform a mean separation on the levels of A.

.... 5. The A*B interaction is not significant. The levels of A are equally spaced numerical values but the levels of B are fixed and qualitative. The experimenter wants to evaluate trends in the levels of factor A.

TECHNIQUE:

A. Either LSD or Tukey's multiple comparison procedure at each level of factor B.

B. Either LSD or Tukey's multiple comparison procedure averaged over the levels of factor B.

C. Fit orthogonal polynomial contrasts at each level of B.

D. Fit orthogonal polynomial contrasts averaged over the levels of B.

E. Nothing new is learned beyond the results of the F-tests from the AOV table.

F. Comparison of means is not appropriate.

G. Scheffe's technique

H. Bonferroni F test for contrasts

I. None of the above are appropriate

PROBLEM #4:

A graduate student has approached you for help in designing her experiment and analyzing the data from the experiment. She has five "treatments" she wants to compare. The treatments are different combinations of setups for desktop computers. For each treatment she will measure a number of different variables, e.g., keying productivity measured in words per minute. Her experimental subjects are employees of a data entry firm. She believes there may be a difference in the response variables for male and female subjects, so will use some of each. She also believes there is a difference in the response variables between younger employees (19-25 years old) and older employees (more than 25 years old) so will use subjects from each of the two age groups. Each subject may be asked to perform the task more than once, i.e., with each of the treatments.

(a.) Suggest an appropriate design for her experiment and give the sources of variation and degrees of freedom for the appropriate ANOVA. You may use a general symbol (e.g., r) for the number of employees or number of employees within a group.

(b.) It is believed that the order in which an employee performs the task (i.e., the order of the treatments for an employee) may have an effect on the response. Would you modify the design you suggested in part a? If so, what would you suggest? Does this place any restriction on the number of employees used? Give the sources of variation and degrees of freedom for the appropriate ANOVA.

(c.) Suppose the treatments are:

A. Desktop PC with 15" monitor with external keyboard and mouse
B. Desktop PC with 17" monitor with external keyboard and mouse
C. Notebook computer alone
D. Notebook computer with monitor blocks and external keyboard and mouse
E. Notebook computer with "Rock Solid" stand and external keyboard and mouse

Would you suggest any analysis (hypotheses tests) other than the basic ANOVA? If so, what analyses?

PROBLEM #5:

For the following two experiments, provide the following information:

1. Type of Randomization, eg, CR, RB, LS, Split-plot, etc;
2. Type of Treatment Structure, eg, single factor, crossed, nested, etc;
3. Identify each of the factors as being fixed or random;
4. Describe the experimental units.
5. An ANOVA Table, Including : Sources of variation, Degrees of freedom, Expected mean squares, Denominator of the F-statistic for testing each elevant effect.

A. A plant scientist wants to investigate plant uptake of nickel under four rates of sludge application. Three varieties of sweet corn were to be studied at each of the four rates of sludge application. For each of the varieties of corn, a total of 40 plants were planted in individual pots. After the plants were established, 10 pots of each variety were randomly selected to receive a rate of sludge. After a period of time, three leaves were randomly selected from each plant and the amount of nickel present in the leaf was determined. In addition to differences in uptake amount due to the different rates of sludge across the three varieties, there was interest in variation among plants treated alike as well as variation among leaves on the same plant.

B. An industrial engineer is investigating the effect of four assembly methods on the assembly time for a color television component. Four operators are randomly selected for the study. Furthermore, the engineer knows that each assembly method produces such fatigue that the time required for the last assembly may be greater than the time required for the first, regardless of the method used in the assembly process. That is, a trend develops in the required assembly time. Each of the four operators uses all four assembly methods, however the order in which the methods of assembly were applied was different for each of the four operators.