METHODS QUALIFYING EXAM

AUGUST 2001

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

In a study of the effectiveness of several new types insecticides, an entomologist designs an experiment to evaluate the relative effectiveness of the five insecticides $(I_1, I_2, I_3, I_4, I_5)$ on brown spotted ticks. From previous studies with similar types of insecticides, she anticipates that under normal conditions, a brown spotted tick exposed to any one of the five insecticides would have a chance of surviving between 30% and 70%.

Five thousand (5,000) very similar brown spotted ticks were partitioned among ten containers, with 500 brown spotted ticks assigned to each container. After an assimilation period, two of the ten containers were randomly selected to receive insecticide I_1, two other containers were randomly selected to receive insecticide I_2, two other containers were randomly selected to receive insecticide I_3, two other containers were randomly selected to receive insecticide I_4, and the final two containers were assigned to receive insecticide I_5. At the end of a 24 hour exposure period, the entomologist examined each of the 5,000 brown spotted ticks and classified the tick as "dead" or "not dead". The response of interest to the entomologist was y_{ij}, the number of "dead" brown spotted ticks (out of the 500) in container j assigned to insecticide, I_i; $i = 1, 2, 3, 4, 5$ and $j = 1, 2$.

(1) Provide an ANOVA table for this experiment, including sources of variation and degrees of freedom. Indicate for each source of variation whether the source is a fixed or random effect. Find the expected mean square for each source of variation in your ANOVA table. Indicate the appropriate denominator of the F statistic for each relevant test.

(2) The customary ANOVA methodology appropriate for this problem provides a test of the null hypothesis that the five insecticides were equally effective. If the test statistic rejects the null hypothesis, explain briefly how you would identify *which* of the five insecticides differed in their effectiveness in controlling brown spotted ticks. Be as specific as possible, and be sure to explain why you belive that your recommended method is appropriate for this problem.

(3) Identify any potential problems with using the type of analysis given in Part (1) and how you would resolve these problems.

(4) After discussing your proposed analysis, the entomologist states, "I think I can just carry out a standard chi-square test of the 5,000 brown spotted ticks classified by a 2x5 contingency table, with the five columns corresponding to the five insecticides and the two rows corresponding to the dead/not dead classification." Do you agree? Explain why or why not.

(5) After the experiment was completed, the entomologist asks, "I plan on conducting out a similar experiment next year on the green spotted tick, again involving 5,000 ticks and five different insecticides. How can I change my experimental design to be more efficient?"

Answer the entomologist's question, taking care to define clearly any notation or special terms you use. In addition, *explain why* you believe your design will be more efficient than the design used in Part (1).

PROBLEM #2

You are given independent observations (Y_{i1}, Y_{i2}), $i = 1, \cdots, N$ on N individuals. Y_{i1} denotes a pre-treatment score and Y_{i2} denotes the post-treatment score on the ith individual, with $E(Y_{ij}) = \tau_j$, $\text{Var}(Y_{ij}) = \sigma^2$, $j = 1, 2$, and $\text{Corr}(Y_{i1}, Y_{i2}) = \rho$ for $i = 1, \cdots, N$. (Assume ρ is known.)

(i) Set this up in a linear model framework, stating all the underlying assumptions and writing explicitly the variance-covariance matrix of the error.

(ii) Again set this up in a linear model framework based on the difference of the responses (Y_{i1}, Y_{i2}), $i = 1, \cdots, N$ rather than the original responses. It means your new response is $Z_i = Y_{i1} - Y_{i2}$, $i = 1, \cdots, N$. Write the new variance-covariance matrix of the error.

(iii) Find the unbiased estimators of the model parameters $\tau_1 - \tau_2$ and σ^2 based on the model in (ii).

(iv) Suppose we assume a bivariate-normal distribution for (Y_{i1}, Y_{i2}), $i = 1, \cdots, N$ then obtain the maximum-likelihood estimators of $\tau_1 - \tau_2$ and σ^2.

(v) Now if you assign uniform prior distributions for $\tau_1 - \tau_2$ and the prior distribution for σ^2 is $\pi(\sigma^2) = 1/\sigma^2$, $\sigma^2 > 0$, then find the conditional posterior distributions of $\tau_1 - \tau_2$ (conditioned on σ^2)

(vi) Hence find the posterior probability that the difference of τ_1 and τ_2 will be positive (conditioned on σ^2).

[Hint: If (X, Y) follows a bivariate normal distribution with means μ_1, μ_2, variances σ^2, σ^2 and correlation coefficient ρ then the density function is

$$f(X,Y) = \left(2\pi\sigma^2\sqrt{1-\rho^2}\right)^{-1} \exp\left\{-\frac{1}{2(1-\rho^2)\sigma^2}\left[(x-\mu_1^2) - 2\rho(x-\mu_1)(y-\mu_2) + (y-\mu_2)^2\right]\right\}]$$

PROBLEM #3

Three different experimental situations are described below. Provide the information requested for each.

A. A study is conducted to examine the effects of three different room temperatures and absence or presence of background noise on the performance of students taking an exam. A total of 48 students are available for the study, with the midterm exam score available for each student. Eight students are to be assigned to each of the six temperature by background noise combinations.

  (i) How would you suggest the students for each temperature by background noise combination be chosen?

  (ii) The score on the exam, taken under the experimental conditions, is recorded for each student. Give the sources and degrees of freedom for the appropriate ANOVA table. Where appropriate, indicate the numerator and denominator mean squares for testing significance of an effect.

  (iii) The exam consists of a multiple choice part, a short answer part, a "work a problem" part, and an essay question. Instead of a single exam score for each student, the scores on the individual parts of the exam are recorded separately. Does this change your ANOVA table? If so, provided the correct ANOVA table and F-ratios.

B. A research assistant to the president of a university collected data for a random sample on n = 100 full time tenure track faculty members. The variables, and their values, are:

SALARY = nine month salary equivalent (in Dollars)

AGE = age, in years

RANK = Academic Rank = 1 for Professor
2 for Associate Professor
3 for Assistant Professor
4 for Instructor, Lecturer, etc.

TENURE = Tenure Status = 0 if not tenured and not tenure track
1 if not tenured, but tenure track
2 if tenured

DEGREE = Final Degree = 1 if bachelors
2 if masters
3 if postmaster, but not doctorate
4 if doctorate

SEX = sex of faculty member = 0 if male
1 if female

TIME = Length of time (in years) since initial appointment to faculty at the university

The president has asked her assistant to determine if there is evidence of discrimination in salaries based on the sex of the faculty member. The research assistant has asked your help.

(i) If you were to fit a model using SALARY as the response variable (the "Y"), which of the other variables would you consider to be class variables (i.e., ANOVA type variables) and which of the variables would you consider to be covariates (i.e., regression type variables)?

(ii) The president asks if there is evidence the change in average salary for each additional year since initial appointment is not the same for males and females. How would you determine if there is evidence of this, i.e., what term or terms would you add to the model?

(iii) Generally, faculty members at the rank of Associate Professor and Professor have tenure, faculty members at the rank of Assistant Professor do not have tenure but are tenure track, and faculty members at the rank of Instructor, Lecturer, etc., do not have tenure and are not tenure track. Could this create any difficulties in fitting a model with all the variables mentioned above included? If so, what is the problem?

C. An analysis of covariance was used to analyze data from an experiment on ten treatments in a Randomized Complete Block Design with five blocks and a covariate, X. Various models were fit to the data. The Error sums of squares (Residual sums of squares) for different models are shown below. The terms in the model are indicated by M = overall mean, BLK = class variable or dummy variables for blocks, TRT = class variable or dummy variables for treatments, and X for the covariate.

Model Includes	SS Error	Error DF
M	393.500	49
M, BLK	332.300	45
M, TRT	321.200	40
M. BLK, TRT	259.000	36
M, BLK, TRT, X	123.120	35
M, BLK, TRT, X(TRT)	82.226	26
M, BLK, TRT, X, X*TRT	82.226	26
M, BLK, X	197.599	44
M, TRT, X	139.928	39
M, X	214.637	48

(i) Give the value of the F-ratio to test if the slopes for the ten lines are equal.

(ii) Assuming a common slope, test that the adjusted treatment means are equal.