M. S. Comprehensive Exam (Part I of Written Exam)
Wednesday, January 3, 1996

You have three hours. Do as many problems as you can. No one is expected to answer all the problems correctly. Partial credit will be given. All problems are worth an equal amount of credit.

Put your solution to each problem on a separate sheet of paper.

Problem 1. (Note: An $F_{\alpha=0.05}$ table is provided for this problem.) In a complete factorial experiment, consider the following mixed-effects model:

$$Y_{ij} \;=\; \mu + \alpha_i + \beta_j + \epsilon_{ij}, \quad i = 1, \ldots, a; \;\; j = 1, \ldots, b,$$

where μ is the overall mean, α_i is a fixed effect corresponding to the ith level of factor A, and β_j is a random effect due to the jth level of factor B. The β_j's are iid $N(0, \sigma_\beta^2)$ variables and the ϵ_{ij}'s are iid $N(0, \sigma_\epsilon^2)$ variables. Furthermore, $\{\beta_j\}$ and $\{\epsilon_{ij}\}$ are assumed to be independent. Let MSA and MSB denote the mean squares for factors A and B, respectively, and let MSE be the mean square for error.

(a) Calculate the expected value of MSA.

(b) Given $a = 3, b = 10, MSA = 5.50, MSB = 7.25$ and $MSE = 0.62$, perform an analysis of variance for the experiment using $\alpha = 0.05$. State your null and alternative hypotheses and summarize your results in a table.

Problem 2. A box contains three balls: one white, one red, and one blue. Balls are successively and randomly drawn from the box, with replacement after each draw. Let N be the number of draws required for each color to appear at least once. Find $P(N = n)$, where $n \geq 3$. (Hint: consider $P(N \geq n)$)

Problem 3. Suppose you want to test three brands of sunscreen (sun-blocking cream).

Design 1 is to select thirty people at random from the target population and apply Brand A to the backs of ten people, B to another ten, and C to another ten, selection of subjects at random. Subjects are given the same number of hours under a tanning lamp and the amount of sunburn is measured.

Design 2 is to select ten people at random from the target population and apply Brands A, B, and C in side-by-side vertical stripes on the back of each person. The response measured in the same way as Design 1.

(a) Give the models and ANOVA tables for designs 1 and 2. Indicate Source, df, and Expected Mean Squares.

(b) How would you select the positions for A, B, and C in Design 2? Explain the advantages of your method.

(c) Which design would you prefer to use? Explain the reasoning for your choice.

Problem 4. Suppose that (X, Y) is uniformly distributed on the triangle:

$$\{(x, y) : x \geq 0, x + |y| \leq 1\}.$$

(a) Find the distribution of Y^2.

(b) Find the conditional density of Y given $X = x$, and $E(Y^2|X = x)$ for $0 \leq x \leq 1$.

(c) Using b), or otherwise, find $E(1 - X)^2$.

Problem 5. Consider the multiple linear regression $Y = X\beta + \epsilon$ where Y $(n \times 1)$, X $(n \times p')$, β $(p' \times 1)$,and ϵ $(n \times 1)$ with $p' = p+1$. Assume that the rank of X is p' $(\leq n)$. Define $H = X(X^TX)^{-1}X^T = (h_{ij})$ as the HAT matrix. Many diagnostic methods have been proposed to evaluate the quality of fit and the quality of prediction in regression.

(a) List <u>five</u> important statistics which are useful for such diagnostic purposes. Give simple expressions for these statistics and briefly describe their use.

(b) It is said that "a subregion in the regression space where prediction is expected to be poor is a subregion in which the fit of the model is expected to be relatively good." True or false? Explain.

Problem 6. Define $Y \sim \mathcal{P}(\lambda)$ to mean that Y has the Poisson distribution with mean λ. Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables with $Y_i \sim \mathcal{P}(e^{\beta_0+\beta_1 x_i})$ where β_0, β_1 are unknown parameters and $x_1, x_2, \ldots, x_n$ are known fixed constants.

(a) Find a two-dimensional sufficient statistic for (β_0, β_1).

(b) The MLE for (β_0, β_1) can be found as the solution of a system of two equations. What are these equations? (Try to write these equations as simply as possible, but do NOT attempt to solve them.)

Problem 7.

(a) An insect lays N eggs, where N has a Poisson distribution with parameter λ. However, the environment is harsh, and the probability of an egg surviving is only p. Assume that the survival of eggs are independent. What is the distribution of X, the number of eggs that survive?

(b) As the environment gets worse, the insect increases the number of eggs it lays by increasing the Poisson parameter λ to go to ∞, while the survival probability of an egg, p decreases to 0 such that $\lambda p \to \mu$. What is the limiting probability that there will be some surviving eggs, i.e., what is the limit of $P\{X > 0\}$?

Problem 8. Let $X_1, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$.

(a) Find the maximum likelihood estimators $\hat{\mu}$, $\hat{\sigma}^2$ of μ, σ^2.

(b) Find $E(\hat{\mu})$.

(c) Find $E(\hat{\sigma}^2)$.

Problem 9. A coin, when tossed, has probability p of coming up heads and probability $q = 1 - p$ of coming up tails. The coin is tossed repeatedly. Determine the expected values of each of the following random variables:

(a) U is the waiting time (number of tosses) until HT appears;

(b) V is the waiting time until HH appears;

(c) W is the waiting time until a string of m heads appears.

Hint: In b) and c) condition on the time of first occurrence of tails to obtain equations that can be solved for $E(V)$ and $E(W)$.