M. S. Comprehensive Exam (Part I of Written Exam)
Sunday, January 6, 2002

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. You have two samples and two measures X and Y on each item in the samples. You plot the two samples on the same scatterplot, using + for one sample and O for the other.

Make three plots which depict the following situations:

(a) X and Y are strongly related, but the correlation is near zero.

(b) X and Y are not correlated within each sample, but, taken as one combined sample, they are correlated.

(c) X and Y are correlated within each sample, but the combined sample shows no correlation.

Problem 2. Answer the following:

(a) Why is it important to have all important predictor variables in a multiple regression model? Compare the consequences of including worthless predictors in the model versus excluding important predictors from the model.

(b) Discuss one aspect of how multicollinearity complicates multiple regression analyses.

(c) Briefly discuss two major problems in the interpretation of the output from typical stepwise multiple regression programs.

Problem 3. An experiment was planned to compare three different fertilizers (A, B, C) on water melon yields. The treatments were randomly assigned according to a Latin square design conducted over a large farm plot which was divided into rows and columns. The watermelon yields (in tons per acre) were recorded below after the growing season.

	Column		
Row	1	2	3
1	9.5 (B)	6.8 (C)	4.9 (A)
2	7.9 (A)	9.1 (B)	6.6 (C)
3	5.6 (C)	7.6 (A)	8.7 (B)

(a) If you were the experimenter, how would you randomize the design? Write an appropriate additive model for the data and state your assumptions on the model.

(b) Conduct an analysis of variance ($\alpha = 0.05$) based on the data and draw your conclusions.

Problem 4. We take four identical marbles. On the first we write the symbols $A_1A_2A_3$, on the second we write A_1, on the third we write A_2, and on the fourth we write A_3. We then put the four balls in an urn and select one at random. Let $B_i, i = 1, 2, 3$ denote the event that the symbol A_i appears on the selected marble.

Are the events B_1, B_2, B_3 mutually independent? Are they pairwise independent? Justify your answers.

Problem 5. Let $X_1, X_2, \ldots, X_n$ be independent and identically distributed (i.i.d.) random variables having the common distribution $F(x) = P(X \le x)$.

(a) Consider an arbitrary parameter $\theta = \theta(F)$. What does it mean for an estimator of θ to be "consistent"?

(b) Suppose F is the Poisson distribution with parameter $\lambda > 0$. Find a consistent estimator for $\lambda e^{-\lambda}$.

(c) Verify that the estimator found in part (b) is indeed consistent.

Problem 6. When a current X (measured in amperes) flows through a resistance Y (measured in ohms), the power generated is given by $W = X^2Y$ (measured in watts). Suppose that X and Y are independent random variables having the probability densities $f_X(x) = 6x(1-x)$ for $0 \le x \le 1$, and $f_Y(y) = 2y$ for $0 \le y \le 1$. Determine the probability density function of W.

Problem 7. Each customer who enters the bookstore will buy a mathematics book with probability p. If the number of customers entering the bookstore is Poisson distributed with mean λ, what is the probability that the bookstore does not sell any mathematics books?

Problem 8. Let $X_1, X_2, \ldots, X_n$ be a sample from a population with the density function (pdf)

$$f(x \mid \theta) = \frac{1}{2} e^{-|x-\theta|}, \quad -\infty < x < \infty, \quad -\infty < \theta < \infty.$$

Find the maximum likelihood estimator (MLE) of θ.

Problem 9. Let $X_1, X_2, \ldots, X_n$ be a sample from a population with the density function (pdf) given by

$$f(x \mid \alpha, \beta, p) = \begin{cases} p\alpha e^{-\alpha x} & \text{for } x \geq 0 \\ (1-p)\beta e^{\beta x} & \text{for } x < 0 \end{cases}$$

where $0 < p < 1$ and $\alpha, \beta > 0$. Find a sufficient statistic for (α, β, p). (Your sufficient statistic should be minimal, but you are **not** required to prove this.)