M. S. Comprehensive Exam
Friday, August 22, 2003

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. Consider the simple regression model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, i = 1, 2, \ldots, n.$$

In ordinary least squares regression, solutions for this problem exist which have many desirable properties when the response variable y is continuous. A typical assumption in this case is that the error terms ε_i are independent, normal random variables with mean 0 and variance $\sigma_i^2 = \sigma^2$ (i.e., constant variance). Suppose that the y_i are **not** continuous, but take on only two values, say 0 or 1.

(a) Show that the errors ε_i in this case are **not** normally distributed if using the simple linear regression model above.

(b) Assume that $E\varepsilon_i = 0$ and $\beta_1 \neq 0$. Find the variance of the error ε_i in terms of $p_i = P(y_i = 1)$ only. Show that the variance is **not** constant for all i.

Problem 2.

(a) Write down a one-way ANOVA model with k number of treatments and n_i number of observations for the i-th treatment, $i = 1, \ldots, k$. Write down the meaning of each symbol. State all assumptions.

(b) For the above model, construct the ANOVA table. Write the formulas for different quantities in the ANOVA table.

(c) Show that the total sum of squares can be written as the sum of treatment sum of squares and error sum of squares.

Problem 3. You are being interviewed by an engineer in an organization in which you want to work. Answer the engineer's questions in everyday English comprehensible to your interviewer who is very bright and has not taken any statistics courses.

(a) I do not understand why you statisticians use techniques that you know contain errors. For example, in the reporting of every poll, a "margin of error" is mentioned. If you know you have made errors in the poll, why don't you correct them? That's what an engineer would do!

(b) I also do not understand all the emphasis you folks put on "simple random sampling." If I want to know how good the paper towels are that we manufacture, I will go to the warehouse and pick cartons from several places and test some towels from each carton. Why is a "simple random sample" better than this? (In your answer you may wish to describe a demonstration using physical objects or computer simulations that would help the engineer to understand you.)

Problem 4.

(a) The c.d.f. (cumulative distribution function) of a random variable X is given by

$$F(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{5} & \text{if } 0 \le x < 1.5 \\ \frac{3}{5} & \text{if } 1.5 \le x < 2 \\ 1 & \text{if } x \ge 2 \end{cases}$$

(i) Find $P(1.5 < X < 4)$.

(ii) Find the expected value of X.

(b) Is the following function $g(\cdot)$ a valid p.d.f. (probability density function)?

$$g(y) = \begin{cases} \frac{60}{7} y(\frac{9}{10} - y) & \text{if } 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

If yes, show that it is a valid p.d.f. If no, explain why not valid.

Problem 5. Let X and Y be independent random variables. Show that

$$P(|X| \le a)P(|Y| > t + a) \le P(|X - Y| > t),$$

for any positive numbers a and t.

Problem 6. Let X_1, X_2, X_3, X_4, and X_5 be independent continuous random variables having a common distribution F and the density function f. Show that the probability of the event $\{X_1 < X_2 > X_3 < X_4 > X_5\}$ does not depend on F.

Problem 7. Suppose $X_i, i = 1, 2, \ldots, n$, are a random sample from a population with density

$$f(x|\theta) = \theta x^{\theta-1},$$

when $0 < x < 1$ and $\theta > 0$.

(a) Find the method of moments estimator of θ. Is it sufficient for θ?

(b) Find the maximum likelihood estimator of $h(\theta) = e^{-1/\theta}$. Is it sufficient for θ?

Problem 8. Show that if $X_1, \ldots, X_n$ are i.i.d. random variables with a p.d.f. $f(x|\theta)$, then the vector of order statistics $X_{(1)}, \ldots, X_{(n)}$ is sufficient for θ.

Problem 9. Suppose θ has p.d.f. $g(\theta) = 2(1 - \theta)$ for $0 < \theta < 1$ (with $g(\theta) = 0$ otherwise) and, conditional on θ, the random variables X_1, X_2, X_3 are i.i.d. Bernoulli(θ). Let A be the event $\{X_1 = X_2 = X_3 = 1\}$.

(a) What is the conditional distribution of θ given A? (Write down the conditional density.)

(b) Find $E(\theta \mid A)$.