M.S. Comprehensive Exam
Friday, January 3, 1992

You have 3 hours. Do as many problems as you can. No one is expected to answer all the problems correctly. Partial credit will be given.

Please put each problem on a separate sheet of paper.

Problem 1: Answer the following questions concerning the design of experiments.

a) What is *confounding* in experimental design and how is confounding used in two-level fractional factorial designs?

b) Suppose I wanted to study four factors in eight runs. Which confounding scheme would be better (say why):

(1) D = ABC, or (2) D=A, or (3) D=AC.

Would your answer change if a previous study had shown that Factors A and C did not interact? Why or why not?

c) Suppose I wanted to study six factors (at two levels each) in 16 runs. Which of the following confounding schemes would I normally prefer and why?

- **i)** Set E = ABC and set F = ABCD.
- **ii)** Set E = ABC and set F = BCD.
- **iii)** Set E = ABD and set F = ACD.

d) Suppose I studied 8 factors A–H in 16 runs. The original factors are A, B, C, and D and E–H are defined by E = ABC, F = ACD, G = ABD, H = BCD. Demonstrate that if I pick only four letters from the above eight that ...

- **i)** sometimes the result is a 2^4 factorial (easy)
- **ii)** sometimes the result is *not* a 2^4 factorial.

Problem 2: A single observation X is to be taken from a mixture of normal distributions that has density

$$f(x|\theta) = \theta(2\pi)^{-1/2}\exp\{-\frac{1}{2}(x-1)^2\} + (1-\theta)(2\pi)^{-1/2}\exp\{-\frac{1}{2}(x+1)^2\}.$$

Find the most powerful test of size $\alpha = .05$ for testing $H_0 : \theta = 1$ against $H_1 : \theta = \frac{1}{2}$. Carefully state any theorem(s) that you use to solve this problem.

Problem 3: Let $x_1, x_2, \ldots, x_n$ be i.i.d random variables with density given by

$$f(x, \mu, \sigma) = \begin{cases} \frac{1}{\sigma} \exp\left\{-\frac{x-\mu}{\sigma}\right\} & \text{if } x \geq \mu \\ 0 & \text{otherwise} \end{cases}$$

where $-\infty < \mu < \infty, \qquad \sigma > 0.$

a) Find a one dimensional sufficient statistic for μ when σ is fixed.

b) Find a one dimensional sufficient statistic for σ when μ is fixed.

c) Find maximum likelihood estimates of μ and σ.

d) Is the maximum likelihood estimate of μ unbiased?

Problem 4: Standardized test scores Y_{ij} are obtained for five individuals who have received a new treatment (experimental group), and for five individuals who have received the standard treatment (control group) . The 95% confidence interval for the difference in mean score between treatment and control groups is $(3.01, 6.59)$, and the conclusion is that the treatment is very effective.
There were also pre-test scores X_{ij} for each participant; however, these were not used in the initial analysis, which was validated by random assignment of subjects to the experimental and control groups. Later the investigator comes to you, the statistician, with the following summary statistics:

Group	$\overline{Y}$	$\hat{\sigma}^2_Y$	$\overline{X}$	$\hat{\sigma}^2_X$	$\hat{\text{cov}}(X,Y)$
Treatment	15.1	2.05	9.7	1.95	.7875
Control	10.3	.95	5.2	.70	.55

The difference in mean improvements between the two groups is
$(\overline{Y_T} - \overline{X_T}) - (\overline{Y_C} - \overline{X_C})$.

a. Derive an estimator of the standard error of this statistic which uses the available summary statistics.

b. Can your estimator in part a. be used to derive a confidence interval for the difference in mean improvement based on the t distribution? If yes, how? If no, why not?

In the remaining parts, suppose you had available all the individual data.

c. Now how would you form a confidence interval for the difference in mean improvement between Treatment and Control groups? What model are you assuming?

d. Name another model which is commonly used in this situation. State the model equation. Is this model a generalization of the model you stated in part c.? What parameter(s) is of primary interest? Can a t-interval be formed?

Problem 5: Let X and Y be independent random variables each of which has a uniform distribution on $[0, 1]$. Define

$$U = \begin{cases} X + Y & \text{if } 0 \le X + Y \le 1, \\ X + Y - 1 & \text{if } 1 < X + Y \le 2. \end{cases}$$

a) Show that U has a uniform distribution on $[0, 1]$.

b) Suppose X still has the uniform distribution on $(0, 1)$, but Y has the density $f(y) = 3y^2$ for $y \in (0, 1)$ and $f(y) = 0$ otherwise. What now is the distribution of U?

Problem 6: Let $f(x)$ and $g(x)$ be density functions. Suppose that $f(x)/g(x) > C$ for all x, where $C > 0$ is a constant. Let X and U be independent random variables, where X has density F and U is uniform on $(0, 1)$. Consider the event A specified by

$$U < \frac{Cg(X)}{f(X)}.$$

a) What is the probability of A?

b) Conditional on A, what is the distribution of X?

Problem 7:
Let X and Y be independent random variables each distributed as $\mathrm{N}(0, \sigma^2)$ with $\sigma^2 > 0$.

Derive the distribution of $W = (X + Y)/|X - Y|$. (You may do this from first principles **or** by quoting standard results in statistics.)