M. S. Comprehensive Exam
Friday, August 23, 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. Answer the following. The parts are unrelated; the solution to (b) does not use (a).

(a) Let (X, Y) have the joint pdf

$$f(x,y) = \begin{cases} 2 \cdot \frac{1}{2\pi} e^{-(x^2+y^2)/2} & \text{if } xy > 0 \\ 0 & \text{otherwise} \end{cases}$$

Find the marginal distributions of X and Y.

(b) Let $S \sim N(0,1)$. Observe S; then toss a fair coin and define T as follows:

$$T = \begin{cases} S & \text{if the toss is "heads",} \\ -S & \text{if the toss is "tails".} \end{cases}$$

Show that T has a normal distribution, but that $S + T$ does not.

Problem 2. A box contains tickets numbered 1 to N. Let X denote the largest number drawn in m random drawings with replacement.

(a) Find $P(X \leq k)$.

(b) Find $P(X = k)$.

Problem 3. Let $X_1, X_2, \ldots, X_n, \ldots$ be independent and identically distributed **continuous** random variables.

(a) Determine the values of $P\{X_1 \geq X_2\}$ and $P\{X_1 \geq X_2 \geq X_3\}$. Also find the value of $P\{X_1 \geq X_2 \geq \cdots \geq X_m\}$ for arbitrary m.

(b) Define a discrete random variable N such that

$$X_1 \geq X_2 \geq \ldots \geq X_{N-1} < X_N \; .$$

That is, N is the point at which the sequence stops decreasing. Use your answer to part (a) to show that $E[N] = e$.

Problem 4. Consider the "regression through the origin" model: $Y_i = \beta X_i + \epsilon_i, i = 1, 2, ..., n$, where the ϵ_i are i.i.d. $N(0, \sigma^2)$.

(a) Find the least squares estimate $\hat{\beta}$ of β and show it is unbiased.

(b) Obtain the standard error of $\hat{\beta}$.

Problem 5. Once upon a time in an age of interplanetary travel and trade, the drug Sarpie became a very important export for the planet of Kibtee. Sarpie was the only known drug which could cure a certain fatal bacterial infection. Sarpie was extracted from the stinger of an insect called bartuck. Like bees on earth, bartucks lived in hives which were populated by one queen bartuck. There were two different methods of keeping the hives and three different strains of bartucks. The following table shows the total Sarpie production for three different hives under the different categories of strains and methods.

Table 1:

	Strain A	Strain B	Strain C	Row Ave.
Method 1	51, 54, 48	55, 54, 59	42, 43, 44	50
Method 2	48, 49, 53	53, 49, 54	43, 39, 38	47.33
Column Ave.	50.5	54	41.5	

(a) Construct the ANOVA table and test to see if interaction effects between methods and strains are present at $\alpha = 0.01$ level.

(b) Test whether or not strain main effects are present at $\alpha = 0.01$ level.

Problem 6. Statistics are widely used and abused. Briefly analyze the following situations. (In each of the questions below, a carefully chosen graph could be used to make your point.)

(a) A sample of 1000 observations from a process **over time** yields a mean of 18.5, with a minimum of 14.0 and a maximum of 24.7.

1. Some will contend that the best prediction of the next observation of this process is 18.5. Under what conditions would you agree with them?
2. What would lead you to predict 14.0?
3. What would lead you to predict 24.7?

(b) Two variables have been measured of 300 people. For this sample, $r = .04$, with a p-value of .29. Some will contend that this is compelling evidence to argue that the two variables are not associated.

1. Under what conditions would you agree with them?
2. Under what conditions would you disagree with them?

(c) X and Y are measured on 470 people. Y is regressed on X, producing the equation $\hat{Y} = 3.09 + .00037X$, and the p-value associated with .00037 is said to be .004. Is it possible that such a small regression coefficient could be significant or has an error been made somewhere? Justify your answer.

Problem 7. Consider the family of pdf's

$$f(x) = \frac{\theta - 1}{x^\theta} \quad \text{for } x \geq 1 \quad (\text{and } f(x) = 0 \text{ otherwise})$$

where $\theta > 1$.

(a) Compute the Fisher information $I(\theta)$ for this family.

(b) Given a sample of size n, find the Cramér-Rao lower bound for the variance of an unbiased estimate of θ.

Problem 8. We define $X \sim \text{Pois}(\lambda)$ to mean that X has a Poisson distribution with pmf $f(x) = \lambda^x e^{-\lambda}/x!$ for $x = 0, 1, 2, \ldots$. Suppose $X_1, X_2, \ldots, X_n$ are independent random variables with $X_i \sim \text{Pois}(e^{\alpha+\beta i})$. Find a complete sufficient statistic for (α, β).

Problem 9. Let X_1, X_2, and X_3 be a sample of size three from a population having probability density function $f(x) = \lambda e^{-\lambda x}$ for $x > 0$ (and $f(x) = 0$ otherwise) where $\lambda > 0$.

(a) Find the maximum likelihood estimator (MLE) for the median of the distribution.

(b) Show that the MLE is unbiased.

(c) Show that the sample median is a biased estimate of the median.

(d) Using an appropriate measure, decide whether the MLE or the sample median is a better estimate for the population median. Justify your answer.