Stat 427 Final Exam Take-Home Portion June 2, 2006

Present your method of solution in a clear, well-labeled manner. Show the details of your calculations; explain and justify your answers fully. Use clear notation, and define the symbols that you introduce. These take-home problems are due *by 10:10am on Thursday, June 8*. Each problem is worth ten points; there are seven take-home problems.

You are to work *completely independently* on all parts of the exam. You are to *discuss no aspect* of the exam with anyone other than me. I will answer only questions of clarification. You may use your class notes, text, and assignments. You may use a computer (statistics packages, spreadsheets, computer algebra systems) as long as you indicate clearly how you use it. You may not consult other aids (such as other books or web sites) unless you get permission from me.

1. I presented students in my STAT 130 course this quarter with a list of 30 letters, and I asked them to memorize as many as possible in 20 seconds. The sequence of letters was the same for all students, but some of the lists came in convenient three-letters groupings (e.g., JFK-CIA-FBI) while others came in less convenient groupings (e.g., JFKC-IAF-BI). I randomly assigned the 38 students in class that day to one of these two groups. My conjecture was that those who received the convenient groupings would tend to score higher than the other group. The data can be found in MemoryLettersS06.mtw, available on our course web site.

Analyze these data by simulation a randomization test. Do one analysis based on the difference in mean results between the groups, and another based on the difference in medians. Produce a histogram of the (empirical) randomization distribution for each analysis, report the (empirical) p-value for each analysis, and summarize your conclusions for each analysis.

2. Can active exercise shorten the time that it takes an infant to learn to walk alone? Researchers randomly allocated twelve one-week-old, male infants from white, middle-class families to one of two treatment groups. Those in the active-exercise group received stimulation of the walking reflexes during four 3-minute sessions each day from the beginning of the second week through the end of the eights week. Those in the other group received no stimulation. The response variable is the age (in months) at which the child first walked alone.

Active exercise group:	9.00, 9.50, 9.50, 9.75, 10.00, 13.00
No exercise group:	9.00, 11.50, 11.50, 12.00, 13.00, 13.25

Conduct an exact randomization test to investigate the research hypothesis, based on the difference in means between the two groups. Report the exact randomization distribution of the difference in means, along with a graph of that distribution. Also report the exact p-value and summarize your conclusion.

3. Simon Newcomb made the first reasonably accurate measurements of the speed of light in 1882. He measured how long it took light to travel from his laboratory on the Potomac River to a mirror at the base of the Washington Monument and back. Based on this travel time, he calculated a measurement of the speed of light. He repeated this process to obtain a total of 66 measurements, which can be found in newcomb.mtw, available on our course web site.
a) Produce and comment on a histogram of these measurements.
b) Calculate the sample mean, sample median, and sample 10% trimmed mean.
c) Use bootstrapping to estimate the standard error of the sample 10% trimmed mean. Produce, submit, and comment on a histogram of the resulting bootstrap distribution of the sample 10% trimmed mean, and report the bootstrap estimate of the standard error.
d) Estimate the population 10% trimmed mean with 95% confidence, using all three bootstrap confidence interval procedures that we have studied.

4. Let X_i represent the number of typographical errors on the i^{th} randomly selected page of a textbook. Suppose that $X_1, \ldots, X_{10}$ are i.i.d. according to a Poisson distribution with parameter λ. Suppose that your prior belief is that $\lambda=.2$ with probability .75 and $\lambda=.5$ with probability .25. You are to observe $X_1, \ldots, X_{10}$ and then make a decision about whether $\lambda=.2$ or $\lambda=.5$ (those are your only two options). If you are correct, you win \$100; if you are wrong you win \$0.
a) Suppose that you find a total of 4 typographical errors on these 10 pages. Determine the posterior expected gain for each action, and identify the posterior Bayes action.
b) Let $X_1 + \ldots + X_{10} = k$. For what values of k is the posterior Bayes action to decide that $\lambda=.5$?

5. Suppose that among three tennis players (call them A, B, and C), one is better than the other two, who are equally good. The best player has probability p (where $p>.5$) of beating the others in a match, and the others have probability .5 of beating each other and probability $(1-p)$ of beating the best player. Your prior probability of being the best is 1/3 for each player. Assume that matches are independent of each other, conditional on p and on which player is the best.
a) Determine the expected value of sample information, as a function of p, in having A play one match against B.
b) Determine the expected value of sample information, as a function of p, in having each player play one match against each other (A plays B once, A plays C once, and B plays C once).

6. Suppose that $X_1, X_2, \ldots, X_n$ are i.i.d. from a contaminated normal distribution, where X_i has a normal distribution with mean θ and standard deviation 1 with probability .9 and X_i has a uniform distribution on the interval $(\theta - k, \theta + k)$ with probability .1, where $k \geq 1$. Consider both the sample mean $\bar{X}$ and sample median $\tilde{X}$ as estimators of the parameter θ.
a) Derive an expression for the ratio MSE($\bar{X}$) / MSE($\tilde{X}$), as a function of k.
b) Produce a graph of this ratio as a function of k.

c) Determine the values of k (if any) for which the sample median has a smaller MSE than the sample mean.

7. Consider using a one-sided, two-sample, unpooled t-test to assess whether the mean recovery time with a new procedure is significantly less than the mean recovery time with a standard procedure.

a) Suppose that the data (in days of recovery) turn out to be:

New procedure:	1, 1, 1, 2, 2, 2
Standard procedure:	3, 3, 3, 4, 4, 4

Write the null and alternative hypotheses, and calculate the test statistic and p-value. Summarize your conclusion.

b) Now suppose that the data (in days of recovery) turn out to be:

New procedure:	100, 100, 100, 200, 200, 200
Standard procedure:	300, 300, 300, 400, 400, 400

Write the null and alternative hypotheses, and calculate the test statistic and p-value. Summarize your conclusion.

c) Prove that multiplying all of the sample values in two groups by the same positive constant (call it c) does not change the value of the two-sample, unpooled t-test statistic.

d) Does multiplying all of the sample values in two groups by the same positive constant (call it c) change the value of the rank sum test statistic? Justify your answer.