Stat 427 Homework Assignment 6 (due Thursday, May 25)
Topics: Randomization Tests, Simulation, Fisher's Exact Test

1. Reconsider the class example involving the age discrimination case filed by Robert Martin. Recall that the data consisted of the ages of the ten employees in Martin's group, three of whom (including Martin, indicated here with an F) had been fired:

25 33 35 38 48 55 (F) 55 (F) 55 56 64 (F)

a) Determine the median age of the three fired employees.
b) Assume that the three fired employees were selected at random from the ten employees. Determine the (exact) probability distribution of the median age among the three fired employees. Also produce a histogram (or other suitable graph) to display this probability distribution.
c) Use this probability distribution to determine the probability that the median age of the fired employees would be at least as large as it was with these data, if in fact the firings had been done at random.
d) How does this p-value compare to the one based on the mean age of the fired employees?

2. Reconsider the previous question about the age discrimination case. We noted in class that knowing the mean age in the fired group suffices to know the difference in mean ages between the two groups.
a) Prove this mathematically.
b) Does a similar result hold for the median: does knowing the median age in the fired group determine the difference in median ages between the two groups? If so, prove it. If not, provide a counter-example.

3. Reconsider the study of sleep deprivation. In class we simulated a randomization test based on the difference in group means.
a) Use simulation (with 9999 repetitions) to approximate the randomization distribution of the difference in group medians. Produce a histogram of this approximate randomization distribution, and summarize its main features.
b) Use your simulation to approximate the p-value of the randomization test, based on the difference in group medians. Also summarize the conclusion that you would draw from this test.
c) Comment on how this analysis based on medians compares to the analysis based on means.

4. Reconsider the experiment that investigated whether a fish oil diet is better than a regular oil diet for reducing blood pressure. Fourteen subjects had been randomly divided into two groups of seven: one group ate a fish oil diet for two weeks, and the other used regular oil. The response variable was the reduction in systolic blood pressure during the study. The data (where a negative value indicates an increase in blood pressure) were:

Fish oil group:	0	0	2	8	10	12	14
Regular oil group:	-6	-4	-3	0	1	2	2

a) Determine the difference in group mean blood pressure reductions.

b) How many ways are there to divide these 14 subjects into two groups of seven each?
c) Determine the exact (one-sided) p-value of the randomization test, by counting how many of these random assignments produce a difference in group means as large (or larger) as you found in a). (You do not need to determine the entire randomization distribution; just count how many randomizations produce such an extreme difference in group means.)
d) Describe the conclusion that you would draw from this test. Also explain the reasoning process by which your conclusion follows.

5. A study conducted in 1996 hoped to show that a new method of coating sutures would lead laboratory rats to develop infarctions at a higher rate than conventional sutures. A total of 14 rats were used in the study, with half randomly assigned to receive a new type of suture and half the conventional suture. It turned out that 10 of the rats developed infarctions. Let the random variable X represent the number of rats in the "new suture" group who developed infarctions.
a) Assuming that the new type of suture was no more/less effective than the conventional suture, and assuming that those 10 rats would have developed infarctions no matter which group they were assigned to, what probability distribution would X have? (Specify the parameters as well as the name of the probability distribution.)
b) Draw a graph of this probability distribution.
c) It turned out that all seven rats in the "new suture" group developed infarctions. Determine the p-value of the test. What conclusion would you draw at the .05 level?
d) Suppose that 6 of the 7 rats in the "new suture" group had developed infarctions. How much would this have changed the p-value?
e) Is there any possible result for this study (still assuming that 10 rats overall developed infarctions) that would have been significant at the .01 level? Explain.
f) Re-answer e), but now allowing for any possible total number of rats who developed infarctions.

6. A study conducted in 2002 investigated whether a nicotine lozenge would help smokers to quit. The experimenters randomly assigned the subjects into one of two groups: 459 smokers were given a nicotine lozenge, while 459 smokers were given a placebo lozenge. The article reported that at the end of 52 weeks, 17.9% of those in the nicotine group had successfully abstained from smoking, compared to 9.6% of those in the placebo group.
a) Create a 2×2 contingency table to display the results of this study.
b) Perform Fisher's exact test to assess the extent to which the data provide evidence that the nicotine lozenge is more effective than a placebo lozenge. Report the p-value, and summarize your conclusion. (Be sure to describe exactly how you perform the relevant probability calculation.)
c) Conduct a chi-square test, and comment on how closely its results match those of Fisher's exact test.