STAT4102 Exam 2, Spring 2007

This is a takehome exam. It should contain your own work only. You are not permitted to talk to or share information with any other person, whether they are in this class or not. You may ask questions of the instructor or the TA. Use of your book and notes is encouraged. Use of other resources (books, webpages, etc.) is unnecessary, but not prohibited.

I certify that this exam is my own work and that I followed the guidelines above.

Name:____________________________________

Signature:____________________________________

Please attach this sheet to the rest of your exam.

Remember to **show all of your work**. Your goal should be to convince the grader that your answer is correct, not merely to write down the correct answer.

Due Date:

This exam will be collected at the beginning of class (2:30 pm) on Friday, April 6. You may also put it in the the instructor's mailbox (Aaron Rendahl) by the Statistics office on the third floor of Ford Hall. Exams will be picked up from the mailbox immediately after class at 3:20 pm. Don't be late.

Class this week:

No lab Tuesday: Brian will have an office hour (Ford 352) from 1:30–3:30.
No class Wednesday: Aaron will have an office hour (Ford 352) from 2:30–3:20.
As usual, Aaron also has an office hour on Monday from 3:30–4:30 and on Friday from 1–2.

I (Aaron) also expect to be in my office (Ford 471) most of the week and available to answer questions either in person or by email.

Scoring:

The exam is out of 50 points. The curve will be no lower than: A-: 42; B-: 34; C-: 26.

1. (10 pts) Suppose you have a data set of 20 data points with sample mean $\bar{x} = 12.24$ and sample standard deviation $s = 5.19$. Upon inspection it looks nearly normally distributed.

   a. (4 pts) Find a 95% confidence interval for the population mean μ. Give an interpretation of this interval in your own words. Does this confidence interval support the hypothesis that $\mu = 10$, or not? Why?

   b. (4 pts) Find a p-value for testing $\mu = 10$ against $\mu > 10$. Give an interpretation of this p-value in your own words.

   c. (2 pts) Do the confidence interval and the p-value lead you to the same conclusion? If so, why would you expect this? If not, why not?

2. (10 pts) Consider again a data set of 20 observations with sample standard deviation $s = 5.19$. You wish to test a null hypothesis of $\mu = 10$ against an alternate of $\mu > 10$.

   a. (4 pts) Against $\mu = 12$, you want the power to be 0.8. Find the rejection region.

   b. (3 pts) What's the size of the Type I error for this rejection region?

   c. (3 pts) Using this rejection region, how large of a sample size would you need to reduce the Type I error to 0.05? Does this increase or decrease the power? Explain.

3. (10 pts) You have one observation X from a Geo(p) distribution. For simplicity, if $X \geq 4$, you will not record the specific value, but only that $X \geq 4$. You wish to test a null hypothesis of $p = 0.9$ against an alternate hypothesis of $p = 0.8$. The distributions follow:

	$P(X = 1)$	$P(X = 2)$	$P(X = 3)$	$P(X \geq 4)$
for $p = 0.9$	0.9	0.09	0.009	0.001
for $p = 0.8$	0.8	0.16	0.032	0.008

   a. (3 pts) List the possible most powerful tests. Calculate the size of the Type I error and the power for each test and sketch their relationship.

   b. (3 pts) Find the most powerful test with $\alpha \leq 0.02$. (It's a randomized test.)

   c. (3 pts) Now suppose you have two observations, and decide to reject the null hypothesis that $p = 0.9$ when at least one of the observations is greater than or equal to 2. What's the size of the Type I error and power of this test? Remember that for independent events A and B, $P(A \cup B) = P(A) + P(B) - P(A)P(B)$.

   d. (1 pt) Would you rather use the test from 3c, which uses two observations but isn't most powerful, or the test from part 3b, which is most powerful, but only uses one observation?

4. (10 pts) Suppose you have a sample of size 30, as follows:

2	6	6	6	6	6	6	7	7	7
7	7	8	8	8	9	10	10	10	10
11	11	12	13	13	14	14	14	15	15

The sample mean $\bar{x}$ is 9.267 and the sample standard deviation s is 3.34.

a. (2 pts) Find the p-value for testing the population mean $\mu = 10$ against $\mu \neq 10$, using a t-test. You can assume the data is near normally distributed.

b. (2 pts) Find the p-value for testing the population median $m = 10$ against $m \neq 10$, using the sign test. See p. 439 for what to do with ties.

c. (4 pts) Now suppose you know the data came from a Poi(λ) distribution, which has the probability function

$$f(x|\lambda) = \frac{\lambda^x}{x!}e^{-\lambda} \quad \text{for} \quad x = 1, 2, 3, \ldots.$$

For testing the population mean $\lambda = 10$ against $\lambda \neq 10$, find the generalized likelihood ratio and use it to find a p-value. Do you expect this value to be similar to the p-value from 4a? Why or why not?

d. (2 pts) Is the likelihood ratio test for these hypotheses uniformly most powerful? Why or why not?

5. (10 pts) To perform a Bayesian analysis, you have calculated the posterior distribution for some parameter θ to be $N(101, 2^2)$.

a. (4 pts) Find a 90% credible interval for θ. How is this interval different than a confidence interval?

b. (3 pts) Find the probability that $\theta > 100$. How is this probability different than a p-value?

c. (3 pts) You wish to decide between the null hypothesis $\theta \leq 100$ and the alternate hypothesis $\theta > 100$. If the cost of a Type I error is 20 and the cost of a Type II error is 10, should you accept or reject the null hypothesis? There is no cost for a correct decision.