QERM 598 - HW 5
Due February 13, 2008
Loveday Conquest and Eli Gurarie

Contingency tables and chi-squared tests

1 Fish scarring

A study was performed analyzing the scarring experienced by adult salmon on their spawning migration in Alaska. The researchers are interested in knowing whether dense fishing activities in Bristol Bay are impacting the fish. The data is counts of scarred and unscarred fish collected at three different periods:

	Period 1	Period 2	Period 3
Scarring YES	14	156	126
Scarring NO	322	188	229

a. Using the notation from the notes, identify r and c and calculate R_i, C_i and N.

b. State a null-hypothesis that explores the effect of time period on fish-scarring.

c. Generate a table of expected frequencies under to the null hypothesis.

d. Obtain a test statistic to test the null hypothesis and identify its null distribution.

e. Perform a test of the null hypothesis based on this statistic.

f. What are your conclusions?

2 Tea Party

David Salsburg wrote an engaging popular history of statistics called *The Lady Tasting Tea: How Statistics Revolutionized Science in the Twentieth Century* (2001, W.H. Freeman & Co., 340 pp). The title refers to a Cambridge University tea party in the early 1900's. One of the guests insisted that tea tastes different depending on whether the tea was poured into the milk or whether the milk was poured into the tea. Most of the people attending the party thought this was nonsense, but R.A. Fisher, who was also in attendance, immediately devised an experiment to test the lady's assertion. The experiment was performed and the lady's assertion was vindicated, thanks in part to a judicious application of the Chi-squared distribution.

a. Can you suggest a very simple experiment that is capable of ruling out the role of random luck in identifying whether a subject can identify the difference in pouring order of milk and tea? How would you perform the test? Be sure to specify the design, the null hypothesis and the test statistic you would use.

b. Let us assume that your subject is, in fact, capable of correctly identifying the difference in 90% of occurences. How many trials of your experiment (at an α level of 0.05) would you need to be 95% confident that you are not falsely failing to reject the null hypothesis? (Note and revel in the quintuple negative so typical of statsy jargon!) What about for a 70% accuracy rate or a 55% accuracy rate? Discuss.

- **Useless hint #1**: The answers are numbers between 1 and 1,000,000.
- **Useless hint #2**: While there is a way to get the answers more or less rigorously by obtaining the distribution under the alternative hypothesis, you can also get a pretty good answer using simulation. Either way is OK, just let me know what you're doing.
- **Comment on jargon**: What you are doing is known as a *power analysis*.