Assignment #5 — Due Friday, October 8, 2010, by 4:00 P.M.

Turn in homework to your TA's mailbox using this sheet as the cover page.

Fill in your name and also circle the *lecture section in which you are registered* and circle the *discussion section you expect to attend* to pick up this assignment.

Name:

Lecture 1 (Larget).	**311:** Tu 1:00 - 2:15pm	**312:** Th 8:00 - 9:15am	**313:** We 1:00 - 2:15pm
Lecture 2 (Hanlon).	**321:** Tu 1:00 - 2:15pm	**322:** We 2:30 - 3:45pm	**323:** We 1:00 - 2:15pm

The following information will help you with problems that ask you to graph contingency table data. Download the file `mosaic.R` that accompanies this assignment. This file contains a function `mosaic()` that modifies the built-in `barchart()` from the `lattice` package in R to create a mosaic plot as described in the text. Load this function into R. (There are many ways to do this, just as you did for `gbinom.R` last assignment: use `source("mosaic.R")`, or `source(file.choose())`, or *Run script...* or *Source File...* from the File menu.)

To enter the matrix of counts, type something similar to this example which uses the data on page 214. The function `matrix()` creates a matrix from the data in the first argument using the second and third arguments to specify the number of rows and columns. The functions `rownames()` and `colnames()` add names to the rows and columns of the matrix which makes the plot more informative. Note that the data is specified column by column, not row by row, in one long array collected with `c()`. You can verify that you set the matrix up correctly by typing its name and seeing it.

```
> fish = matrix(c(1,49,10,35,37,9),nrow=2,ncol=3)
> rownames(fish) = c("Eaten","Not eaten")
> colnames(fish) = c("Uninfected","Lightly infected","Highly infected")
> fish
> mosaic(fish)
```

In the example, the plot shows the relative frequencies within each column. If you want to do the same for rows, use `t()` to transform the matrix.

```
> mosaic(t(fish))
```

Please answer the following questions.

1. The blood type distribution in the United States during World War II was thought to be the following: type A 41%, type B 9%, type AB 4%, and type O 46%. It is estimated that 4% of soldiers with true blood type O blood were misclassified as type A; 88% of those with type A blood were correctly typed; 4% with true type B were typed as A; and 10% with true type AB were typed as A. A solider was wounded and brought to surgery. He was typed as having type A blood. What is the probability that this was his true blood type?

2. Use the data set from p.227, Problem 11. But answer these questions.
   (a) Examine a mosaic plot that compares the estimated conditional probabilities of success of wart removal given the treatment. Include this plot with your solution.
   (b) Find a point estimate and a 95% confidence interval for the difference in the success probabilities of the two therapies. Interpret the result.
   (c) Find a point estimate and a 95% confidence interval for the odds ratio of the two therapies. Interpret the result.

3. Use the data set from p.229-230, Problem 20. But answer these questions.

   When conducting a hypothesis test, make sure to include the following: statements of the null and alternative hypotheses, the observed value of the test statistic, the p-value, and an interpretation of the result. To compute p-values based on a χ^2 distribution, use the R function `pchisq.()`, which computes a the probability that a χ^2 random variable is *less than* a given value. As p-values here are the probability that a χ^2 random variable is *more than* the test statistic, you would use a command as in the following example. The p-value for the G-Test with observed test statistics $G = 8$ and $df = 2$, is given by

```
> 1-pchisq(8,df=2)
```

   (a) Display the data with a mosaic plot to highlight proportions of people in each smoking class within each diet class.
   (b) Perform the χ^2 test of independence on the data set.
   (c) Perform the G-test on the data set.
   (d) Relate the results of these hypothesis tests to what you saw in the mosaic plot. What features apparent in the plot do you think had the largest influence on the results of the tests?
   (e) The test statistics for the two tests involve summing over all cells in the table. Which cells contributed the most to these sums? How does this relate to your answer of the previous part of the question?

4. Textbook p.229, Problem 19.