Categorical Data Analysis: HW 5

1. Refer to the pdf about safety information of the drug "Zelnorm" available on our course webpage. Patients were either given the drug or a placebo and the number of cardiovascular ischemic events (i.e., heart attacks) were reported in each group. See if you can find out which test they used that resulted in the stated P-value of 0.024. State the null an alternative hypothesis for this test.

2. The slides on Fisher's exact test mention that we can, among others, order the tables (to find out which are the extreme ones) by the observed X^2 statistic for each table. This might give a slightly different ordering of the tables (i.e. which tables are considered extreme) in the 2×2 case. An order based on table odds ratios (or the first cell count) need not be equivalent to an order based on the value of X^2 for a given table. However, when extending Fisher's exact test to the general case of $I \times J$ tables, we need a different criterion since the odds ratio only applies to 2×2 tables.

   Consider the example from the R handout (on our course webpage) on testing independence between eye color and hair color of students. We can look at all possible tables with the given row and column margins (how many are there?), and for each table compute the X^2 statistic. Next, we can figure out the null probability for each table. This is based on the *multivariate* hypergeometric distribution, the formula of which we derived in class. Theoretically, you can compute this for each possible table and then get the P-value by summing the null probabilities for all tables that have X^2 value as large or larger than the observed one. (Remember, the larger X^2, the more evidence for an association and against independence.) If you want, you can do this, but it might be computationally complex because of the number of possible tables with the given margins. *An alternative*, similar to what we did in the one-variable case, is to randomly sample (let's say 100,000) tables with the given margins, for each computing X^2 and then just getting the proportion of tables that have X^2 as large or larger than the observed X^2. This is exactly what is done when you call `chisq.test()` in R with the `simulate.p.value=TRUE, B=100000` option. Use it to get the exact P-value for testing independence between eye color and hair color based on the data in the handout. Compare this exact P-value to the asymptotic P-value based on the Chi-square approximation of X^2.

3. The file "Relief of Dysmenorrhea" on our website refers to a dataset that compares placebo (treatment A) with a low-dose analgesic (treatment B) and high-dose analgesic (treatment C) for relief of primary dysmenorrhea. Subjects in the study were divided randomly into six groups, the possible sequences for administering the treatments. Each subject was given each treatment. At the end of each treatment period, the subjects rated the treatment as giving no relief (0) or some relief (1). Fit a model to these data, using GEE, that has a sequence effect and a treatment effect, but assume a common treatment effect for each sequence, i.e., no treatment by sequence interaction. Does it matter in which sequence the treatments are administered? Also, use this model to come up with an ordering of the treatments based on their effectiveness.