Categorical Data Analysis: HW 3

1. An advertisement by the pharmaceutical company Schering Corp. for the allergy drug Claritin (which I use!) mentioned that in a pediatric randomized clinical trial, symptoms of nervousness were shown by 4 of 188 patients on loratadine (=Claritin).

   (a) Find the exact confidence interval for the true proportion of patients showing symptoms of nervousness and compare it to the score confidence interval.

   (b) For the same number of patients (188), it was also recorded how many adverse events they experiences during a three week period. 175 patients experienced no adverse event, 5 patients experienced exactly one adverse event, 2 patients experienced exactly two adverse events and the rest experienced more than two adverse events. Test the hypothesis that the true proportion of patients experiencing no, one, two or more than two adverse events is equal to 98%, 1%, 0.5% and 0.5%, respectively. Use i.) an asymptotic test and ii.) an exact test. For the exact test, you will probably have memory issues when you trying to create all possible tables using my naive way with expand.grid(). (Try it, but make sure you save your work before!). You can use a more sophisticated method to create all possible tables with R, i.e., you can write your own code that does that. There are 1,143,135 distinct possible tables, which is not a large number and one that R can handle. Also, I think that there are R packages that can do such a thing, maybe the permute package? **Alternatively**, you can run a Monte Carlo test by simulating many tables under the null hypothesis. Actually, this option is implemented in R's `chisq.test()` function, and all you really have to do is to invoke it using `simulate.p.value=TRUE, B=10000`, see the help file.

2. This is an important example from Genetics: Genotypes AA, Aa, and aa are believed to occur with probabilities $\theta^2, 2\theta(1-\theta)$ and $(1-\theta)^2$. A multinomial sample of size n has frequencies (n_1, n_2, n_3) of these three genotypes.

   (a) Find the MLE of θ.

   (b) Find the asymptotic standard error of the MLE.

   (c) Give the results of a test that tests this belief when $(n_1, n_2, n_3) = (4, 8, 15)$.

3. There is a rumor that females are better at math than males. Let's collect data which we can use to test this rumor. Because the opinion might depend on gender, you want to sample both females and males. So, the two variables involved are gender and opinion on whether females are better at math.

   (a) In context, describe a sampling strategy that would lead to Poisson sampling.

(b) In context, describe a sampling strategy that would lead to multinomial sampling.

(c) In context, describe a sampling strategy that would lead to product multinomial (product binomial) sampling.