Week 7 Chi-Square Clinic

Instructions: There is a dataset built into R called HairEyeColor that contains a contingency table for n=592 statistics students. Four variants of hair color are crossed with four variants of eye color with cell counts of the number of people fitting each description. Your goal is to conduct a thorough chi-square analysis using both conventional and Bayesian techniques and to write an interpretation that integrates the statistical results.

1. First, just for simplicity, combine the males and females into one 4x4 contingency table:  
     
   HEcombined <- HairEyeColor[,,1] + HairEyeColor[,,2]  
     
   Note the double commas in the selector: HairEyeColor is a 3D table! Review the result by typing HEcombined at the console. Check the total number of observations in the new dataset with sum(HEcombined). Paste the results below:
2. Review the absolute proportion of observations in each cell with:  
     
   HEcombined/sum(HEcombined)  
     
   Paste the results below. Review the results either row by row or column by column.   
     
     
     
   Later, we are going to focus on a subset of four cells where the real action is located. Can you pick out which ones they might be?
3. Calculate the chi-square value and test the significance of the result. You can use the chisq.test() procedure to accomplish this:  
     
   chiOut <- chisq.test(HEcombined)  
   chiOut  
     
   Paste the results below. Report the results of the significance test. Why are there nine degrees of freedom? What is the null hypothesis? Is the null hypothesis rejected?
4. Examine the “residuals” from the chi-square test with chiOut$residuals. Paste the table of residuals below. Residuals represent how far an observed value was from the expected value. A large positive residual means that the observation for a cell was much higher than expected. A large negative residual means that the observation for a cell was much lower than expected. Large residuals (negative or positive) indicate the cells that made the most powerful contribution to the value of chi-square. Therefore, cells with large residuals show where the “action” is with respect to non-independence.
5. Based on the residuals, choose two rows and two columns that have the most interesting stuff happening in them. For example, I chose rows 1 and 4 and columns 1 and 2 with the following command:  
     
   HEsmall <- HEcombined[c(1,4), c(1,2)]  
     
   Verify that you got what you wanted by typing HEsmall at the command line. Paste the resulting 2x2 table below.
6. Now conduct a Bayesian contingency table analysis with the following line of code:  
     
   ctOut <- contingencyTableBF(HEsmall, sampleType="poisson", posterior=TRUE, iterations=10000)  
     
   Review the results with summary(ctOut) and paste the results below.
7. Examine histograms of the posterior distributions of the proportions of different colored eyes. Here’s an example of code that will plot the ratios formed by the entries in the first row:  
     
   brownBlueRow1 <- ctOut[,"lambda[1,1]"] / ctOut[,"lambda[1,2]"]

hist(brownBlueRow1)   
  
Modify that code to also figure the proportions for the second row. Describe what you are seeing and make sense out of it. You might want to refer back to HEsmall to remember what the original raw data was in each cell.

1. Finally, display a histogram of the difference in proportions between the two rows. Explain what you see. Just based on a visual inspection, does the HDI overlap with zero? Is there a credible difference between the two rows with respect to the proportions of different colored eyes?