**IST 772 Week 8 Breakout: Analyze Independence for Categorical Associations**

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.

1. First, 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.

1. Review the proportion of total observations in each cell with:  
   HEcombined/sum(HEcombined)  
   **Review the results going row by row or column by column. Add comments to describe what you see.**

# From what we see from proportion of total observations, students with black hair tend to have brown eyes in 11% of the cases followed by blue eyes, hazel and then green. For people with brown hair, 20% of the students in observation tend to have brown eyes, followed by blue, hazel and green. For Red hair, 4% of the students have brown eyes, followed by blue, and hazel and green with the same proportions. For blond hair, 15% of the students have blonde hair, followed by green eyes, hazel and brown in the end.

1. Calculate the chi-square value and test the overall significance of the test of independence. You can use the chisq.test() procedure to accomplish this.  
   chiOut <- chisq.test(HEcombined)  
   chiOut

**What are the results of the significance test?**

The test indicates that there is a significant association between the variables Hair and Eye (HEcombined), with a very low p-value that is smaller than the usual significance level of 0.05.

**Why are there nine degrees of freedom?**

There are nine degrees of freedom because the contingency table formed by the cross-classification of Hair and Eye has four rows (for Hair categories: Black, Brown, Red, Blond) and four columns (for Eye categories: Brown, Blue, Green, Hazel). Thus, the number of degrees of freedom can be calculated as df = (number of rows - 1) x (number of columns - 1) = (4-1) x (4-1) = 9.

**What is the null hypothesis?**

The null hypothesis in this case is that there is no association between the variables Hair and Eye. That is, the distribution of hair colors is the same for all eye colors, and vice versa.

**Is the null hypothesis rejected?**

Yes, the null hypothesis is rejected based on the very low p-value (smaller than the usual significance level of 0.05), and we can conclude that there is a significant association between the Hair and Eye variables.

1. Examine the “residuals” from the chi-square test with chiOut$residuals. 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.** **Where does this occur?**

The largest positive result is 7.05 which corresponds to blond hair with blue eyes.

The largest negative result is -5.05 which corresponds to blond hair with brown eyes.

1. Make sure to library BayesFactor so you are able to conduct a Bayesian contingency table analysis with the following line of code.

ctOut <- contingencyTableBF(HEcombined, sampleType="poisson", posterior=FALSE, iterations=10000)

**Review the results with ctOut.** **What does this odds ratio tell you?**

The Bayes factor is 3.318262e+26, which means that the data is overwhelmingly more likely under the alternative hypothesis than under the null hypothesis that there is a relation between the hair color and eye color.

1. Write a statement regarding the relationship between hair color and eye color using evidence from both the frequentist and Bayesian analyses.