Your Name: _____

Names of people you worked with: _____

Instructions: Work on this problem in class with your group. Do your best. This piece of paper will be collected during class.

Task: Once again, consider a setting where we are interested in comparing the soda preference across three of the Claremont Colleges. The following table describes the soda preferences for a sample of Pomona, Pitzer, and CMC students.

College	Soda Pepsi	Preference Root beer	Sprite	Total
CMC	10	17	9	36
Pitzer	6	16	2	24
Pomona	24	27	9	60
Total	40	60	20	120

Describe a physical object you could use to run a complete hypothesis test on these observations if you had no knowledge of the mathematical theory which describes a χ^2 test.

1. What is the object? Describe any relevant characteristics of the object.

2. What do you do with the object to make sure the null hypothesis is true? (Why do you need to make sure the null hypothesis is true?)

3. After insuring the null hypothesis is true, what do you calculate?

4. Describe the tedious part.

5. After a while, you are ready to make a conclusion. What is the process by which you make your conclusion?

Solution:

1. What is the object? Describe any relevant characteristics of the object.

   - The object is a set of cards.
   - There should be 120 cards.
   - 40 cards will be blue (the Pepsi cards); 60 cards will be brown (the Root beer cards); 20 cards will be green (the Sprite cards).

2. What do you do with the object to make sure the null hypothesis is true? (Why do you need to make sure the null hypothesis is true?)

   - Shuffle the cards.
   - Deal the cards into three groups: 36 cards to CMC; 24 cards to Pitzer; 60 cards to Pomona.
   - The null hypothesis must be true so that we can understand the null sampling distribution. That is, how does the statistic vary from sample to sample just by random chance. Even when there are no inherent preferences, there is still always variability which we can quantify.

3. After insuring the null hypothesis is true, what do you calculate?

   - Calculate the chi-square test statistic, as in WU 15.

$$X^2 = \sum_{\text{all cells}} \frac{(Observed - Expected)^2}{Expected}$$

4. Describe the tedious part.

   - Repeat steps 2 & 3 many many times.

5. After a while, you are ready to make a conclusion. What is the process by which you make your conclusion?

   - Count the number of null X^2 values which are larger than the observed X^2 value.
   - The proportion of null X^2 values which are larger than the observed X^2 value estimates the p-value.
   - The p-value is the probability of the observed data or more extreme given the null hypothesis is true.
   - If the p-value is small, reject the null hypothesis and claim that there is an association between soda preference and school.