Chi Square (X^2) Modeling Using Candy

The Chi Square test is often used in science to determine if data you observe from an experiment is close enough to the predicted data. In genetics, for instance, you might expect to get a 75% to 25% ratio if you crossed two heterozygous tall plants (Tt x Tt). Calculating the X^2 values help you determine whether the results follow the prediction and if the variations from the exact ratio are due to random chance. It's the question of "how close is close enough?" If the numbers differ greatly from your expected results, then it's possible that other factors may be influencing your results.

The Chi-square test is intended to test how likely it is that an observed distribution is due to chance. It is also called a **"goodness of fit"** statistic, because it measures how well the observed distribution of data fits with the distribution that is expected if the variables are independent.

Another way to describe the Chi-square test is that it tests the **null hypothesis** that the variables are independent, that there is no relationship between the two things being tested. Wherever the observed data doesn't fit the model, the likelihood that the variables are dependent becomes stronger, thus proving the null hypothesis incorrect!

Question: Do Companies Selling Candy Equally Distribute Candy Colors in Bags?

Null Hypothesis: Candy is evenly distributed; each bag contains the same number of colors.
Alternate Hypothesis: One (or more) colors is found in greater frequency.

Materials: several bags of colored candy, such as M & M's, Skittles, Reese's Pieces, or Gummy Bears.

Procedure:

1) Look into the bag and determine how many colors are present and write them into Table 1
2) Without counting, estimate the number (percentage out of 100%) of each if they are all evenly sorted to packages. This is your "Percentage Estimate."
3) Sort the candy and write down the number of each color into Table 1 under "Number Observed"
4) Complete the table by determining the total number of candies and "number expected" columns

Color of Candy	Percentage Estimate	Number Observed	Number Expected (total # of candy x % estimate)
		Total Number of Candies =	

As you look at the data above, consider the two comparable numbers. The number you would expect to count if your percentage estimate was correct, and then the number you actually counted (number observed). For example, if you initially thought that you'd see 25% yellow candies, and you counted 200 pieces, you would then expect to see 50 yellow candies. You may have only counted 40 yellows.

The Chi Square (X^2) Equation

$$X^2 = \sum \frac{(o-e)^2}{e}$$

In order to complete the calculation, you sum each of the traits (colors) that you measured. To help you with this, we will break the process into steps.

	Classes (colors)	Expected (e)	Observed (o)	$\frac{(o-e)^2}{e}$
1				
2				
3				
4				
5				
	Sum (add the values from row 1-5); this is your X^2 value			

Use the chi square chart below to determine if your X^2 supports or rejects your hypothesis.
The degrees of freedom is determined by subtracting 1 from the number of colors you analyzed. (For example, if you had 4 colors to count, the degrees of freedom is 3)

Degrees of Freedom	good fit					poor fit		
	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01
1	—	0.001	0.004	0.016	2.706	3.841	5.024	6.635
2	0.020	0.051	0.103	0.211	4.605	5.991	7.378	9.210
3	0.115	0.216	0.352	0.584	6.251	7.815	9.348	11.345
4	0.297	0.484	0.711	1.064	7.779	9.488	11.143	13.277
5	0.554	0.831	1.145	1.610	9.236	11.071	12.833	15.086

Summary and Analysis

1) What was your initial hypothesis?

2) How do you show that your hypothesis is correct (or incorrect)?

3) Explain what is meant by a "good fit"?

4) Propose a way that a chi square analysis could be used in other experiments, such as genetics or drug trials.