Chi Square Analysis Lab

12/11/14 (Lab #7)

Introduction

When crossing individuals, there is sometimes deviation from the expected results. A chi-square (x^2) analysis can help to determine whether this is due to chance or another reason.

A null hypothesis, important while conducting a chi-square analysis, is a prediction about whether the observed pattern of data and the expected pattern of data are effectively one and the same. The null hypothesis for a chi-square analysis is always the same one:

Any difference between observed and expected data is due to chance.

This hypothesis will be confirmed or refuted in the experiment once a value is calculated and compared to a table of critical values. If the value is smaller than the critical value, the null hypothesis will be accepted and if it is larger, the null hypothesis will be rejected.

The formula for the chi-square analysis consists of O, the observed number (actual count), E, the expected number, Σ, indicating the sum of the results for each category. This is the formula:

$$x^2 = \frac{\Sigma (O-E)^2}{E}$$

The value of x^2 increases as the difference between observed and expected values increases.

This investigation will allow students to practice using the chi-square test with familiar objects, such as M&Ms. Students will write and test a null hypothesis, determine degrees of freedom for an investigation, calculate the actual x^2 value, and determine if the null hypothesis should be confirmed or refuted based on the calculations.

Hypothesis

Any difference between the observed and expected data is due to chance.

Materials

- bag of M&Ms
- paper plate
- pen and paper
- clean hands

Procedure

1. Wash hands with soap and water.
2. Gather materials.
3. Open the bag of M&Ms and pour onto plate. Do Not Eat Yet!
4. Seperate the candies into color categories.
5. Count the number of candies in each category.
6. Record results in Data Table 1.
7. Calculate the expected number of each color and record values in Data Table 1.
8. Complete the calculations indicated in the remaining rows of Data Table 1 to determine the Chi-Square value for the data.

Data

Table 1 - Results for one bag of Milk Chocolate M&Ms

	Brown	Blue	Orange	Green	Red	Yellow	Total
Observed (O)	46	57	46	46 ~~20~~	20 ~~56~~	56	271
Expected (E)	35.23	65.04	54.2	43.36	35.23	37.94	271
Difference (O-E)	10.77	8.04	8.2	2.64	15.23	18.06	
Difference^2 $(O-E)^2$	116	64.64	67.2	6.97	231.95	326.16	
$(O-E)^2/E$	3.29	0.994	1.24	0.16	6.58	8.6	
$X^2 = \Sigma (O-E)^2/E$							20.86

Table 2 - Collected Results from the Class (6 bags)

	Brown	Blue	Orange	Green	Red	Yellow	Total
Observed (O)	293	275	302	286	186	281	1623
Expected (E)	211	389.5	324.6	259.7	211	227.2	1623
$(O-E)$	82	114.5	22.6	26.3	25	53.8	
$(O-E)^2$	6724	13110.25	510.76	691.7	625	2894.4	
$(O-E)^2/E$	31.87	33.66	1.57	2.66	2.96	12.74	
$X^2 = \Sigma (O-E)^2/E$							85.4

Error Analysis

Mistakes may have been made in counting the exact number of candies for each color category. Calculation errors are a possibility as well.

continued on page 41

Conclusion

The hypothesis for this investigation was that any difference between observed and expected data should purely be due to chance. Based on the obtained results and calculations, this null hypothesis should be rejected. The calculations showed that the chi-square value was much larger than the critical value, which means that the disparities between observed and expected values must be due to ~~oth~~ factors other than chance.

The critical value was chosen based on the degrees of freedom (number of categories − 1), which was five for the M&Ms since there were six color categories. The chi-square value was calculated by squaring the difference of the observed and expected values, dividing that by the expected value, and adding the results for each separate category.

Students counted the candies in one bag first and recorded the data. The chi-square value was calculated and compared to the critical value (11.07). The chi-square value was 20.86, which led to the rejection of the null hypothesis. Then, the results from each group were combined and the calculations were performed once more, only with six bags of milk chocolate M&Ms this time. The chi-square value came out to be 85.4, which led to another rejection of the null hypothesis, because this value was much larger than the critical value.

Based on these results, students could conclude that the differences between observed and expected data must be due to other factors, not just chance. Possible explanations for this could be that the expected data from the Mars Company was outdated and did not apply anymore. Also, the company might not have made sure that the correct (expected) number of candies of each color got into each bag. The methods of production might not have ensured that each bag received the percentage of different colors shown on the M&M website. Therefore, the company was not true to the word on its website and didn't regulate the number of different colored candies in each bag. Data was collected from the entire group to obtain a more accurate and viable result, since a larger amount of data will provide a more correct conclusion.

In this experiment, students learned to calculate the chi-square value from observed data and compare it to the critical value based on the degrees of freedom in order to determine whether the differences in the observed and expected data were due to chance or not. Students applied this statistical test to familiar and simple objects, such as M&Ms, in order to gain a better understanding of the ~~poth~~ scientific concept.