Chi-Square Analysis Using M&M's

Introduction

Do you ever wonder why the package of M&M's never seems to contain enough of your favorite color? Or why does it seem that there is more of one color than another? Is the percentage of each color really different from one package to another? These questions can be answered by using a Chi-square analysis.

Concepts

- Goodness-of-fit test
- Chi-square distribution
- Probability

Background

Chi-square analysis is used to perform hypothesis testing on nominal and ordinal data. *Nominal data* measures qualitative data assigned to predetermined categories. For example, the gender of the respondent in which the categories are male and female. *Ordinal data* has all the properties of nominal data with the added ability to rank order of the values from highest to lowest. An example being a restaurant can be rated five stars, four stars, three stars, etc.

One way to conduct the Chi-square analysis with M&M's is to perform a *goodness-of-fit* test. This test uses a sample to test whether a frequency distribution fits the predicted distribution. This type of statistical analysis allows us to compare the observed (O) and expected (E) frequencies occur by chance and if they are statistically significant. In this example the observed category is counted from a sample set (a bag of M&M's) and the expected counts are calculated based on the percentages provided by the manufacturer and the total of M&M's in the sample. The expected percentages of each color per bag are shown in Table 1.

Color	Percentage
Blue	24%
Orange	20%
Green	16%
Yellow	14%
Red	13%
Brown	13%

Table 1. Percentages of each color provided by the Mars Co.

For example, if a bag contains 500 M&M's the expected number of green candies would be 80 (500 total in the sample × 16%). By comparing the observed and expected values a statistic known as the *chi square* or X^2 can be determined. This value will be compared against a data table to determine the probability of obtaining a particular chi-square value. The probability value explains what the chances are that the differences in the observed vs. expected data are due to chance (such as a sample error). In other words, the value tells us whether or not the deviation from the expected ratios is significant.

To test a hypothesis the X^2 value must be calculated using equation 1. It is important to note with this formula, when all else is equal, the chi-square value increases as the difference between the observed and expected values increase.

$$X^2 = \sum \frac{(O-E)^2}{E}$$ *Equation 1*

Publication No. 11151
061616

Degrees of Freedom	Accept the null hypothesis				Reject the null hypothesis	
	Probability					
	0.9	0.5	0.25	0.1	0.05	0.01
1	0.016	0.46	1.32	2.71	3.84	6.64
2	0.21	1.39	2.77	4.61	5.99	9.21
3	0.58	2.37	4.11	6.25	7.82	11.35
4	1.06	3.36	5.39	7.78	9.49	13.28
5	1.61	4.35	6.63	9.24	11.07	15.09

Safety Precautions

The materials used in this activity are considered nonhazardous. However, this activity should not be performed in a laboratory where chemicals are used. Once food grade items are brought into the lab they are considered chemicals and should not be consumed.

Materials

M&M's
Calculator
Paper
Paper towel
Pen

Procedure

1. Wash hands thoroughly before beginning this activity.
2. Spread clean paper towels out on the desk where the candy will be handled.
3. Separate the M&M's into color categories.
4. Count and record the number of M&M's in each category.
5. Once the entire class has counted their candy compile the class data for each color category
6. Determine the total number of M&M's counted in the class.
7. Calculate the expected numbers of M&M's in each color category using the information provided in the background section.
8. Record both the expected and observed counts of M&M's on the worksheet.
9. Determine the probability that the difference between the observed and expected values occurred by chance or a sign of something else.

Connecting to the National Standards

This laboratory activity relates to the following National Science Education Standards (1996):

Unifying Concepts and Processes: Grades K–12
Evidence, models, and explanation
Constancy, change, and measurement
Content Standards: Grades 5–8
Content Standard G: History and Nature of Science; nature of science
Content Standards: Grades 9–12
Content Standard G: History and Nature of Science, nature of scientific knowledge

Tips

- Results of this experiment often vary from the percentages expected based on data from Mars Chocolate North America. M&M's are made in large production batches which are blended and mixed throroughly. However, individual packages are filled by weight, not by count, on high speed equipment unusual color distribution may occur.
- Allow students to be creative with their candy choices. Oddly enough, different types of M&M's have different expected color distribution. The expected color percentages of each M&M type are listed below in Table 3.

	Color						
M&M Variety	**Blue**	**Orange**	**Green**	**Yellow**	**Red**	**Brown**	**White**
Peanut	23%	23%	15%	15%	12%	12%	
Kid's Minis	25%	25%	12%	13%	12%	13%	
Dark	17%	16%	16%	17%	17%	17%	
Peanut Butter & Almond	20%	20%	20%	20%	10%	10%	
Pretzel	28.5%	14.3%	14.3%	14.3%	14.3%	14.3%	
Coconut			25%			38%	37.5%

Table 3.

Sample Data

Count the total number of M&M's in the sample population and determines the expected frequencies of each color. If there are **632** M&M's in the bag then the expected number of each color are as follows:

Colors	**Expected Percentage**	**Expected Frequency**
Blue	*0.24*	*151.68*
Orange	*0.2*	*126.4*
Green	*0.16*	*101.12*
Yellow	*0.14*	*88.48*
Red	*0.13*	*82.16*
Brown	*0.13*	*82.16*

Use the formula $(O–E)^2/E$ to calculate the chi-square value. Notice the value of 21.53 is the sum of the chi-square value for each color.

The Calculated Chi-Square					
Rating	O	E	(OE)	$(O–E)^2$	$(O–E)^2/E$
Blue	*155*	*151.7*	*3.3*	*10.89*	*0.07*
Orange	*161*	*126.4*	*34.6*	*1197.2*	*9.47*
Green	*90*	*101.12*	*–11.12*	*123.7*	*1.22*
Yellow	*61*	*88.48*	*–27.48*	*755.2*	*8.54*
Red	*92*	*82.16*	*9.84*	*98.8*	*1.20*
Brown	*73*	*82.16*	*–9.16*	*83.9*	*1.02*
				Total chi-square value	*21.53*

In order to determine if the chi-square value is within the acceptable range it is compared against the value in the 0.05 column (statistical validity) where it intersects with 5 degrees of freedom (number of categories, 6 minus 1). This value is 11.07. Since the chi-square value is 21.53, greater than 11.07 these changes are not due to random variances and the hypothesis would be rejected.

Acknowledgements

Special thanks to Kathy Van Hoeck, York High School, Elmhurst, IL for sharing this activity with Flinn Scientific.

Thank you to Mars Chocolate North America for providing expected color percentages in each type of M&M candy.