Statistics Research Project

Derrik Echevarria

April 27, 2017

**M&Ms Colors in Fun Sized Bags**

Statement of the Problem

Are all colors created equal? I was curious to know whether or not, if I open a fun sized bag of M&Ms, will I find the same number of candies in each color? M&Ms come in the colors red, blue, green, orange, yellow and brown. I expect that each bag is filled at random, so that each color will not be represented equally. I decided to test this.

Methods

The total population for this research would be every fun sized bag of M&Ms. Being that it is not feasible to access and count every bag of fun sized M&Ms that currently exists, I chose a much smaller sample. I went to the grocery store and procured one 10.5 Oz bag of fun sized M&Ms which contained 21 fun sized bags of M&Ms.

To find my counts, I simply opened each of my 21 fun sized bags of M&Ms separately, and noted how many of each color were in each bag. I recorded the results of each bag individually. I also made note of the totals of all bags.

The data was comprised of counts of different colors of M&M candies. I believed that these counts would be normally distributed about the true mean for each color, satisfying the Central Limit Theorem.

The total number of fun sized bags I counted for my sample was 21. This adequately satisfies the Law of Large Numbers, particularly because I had no reason to believe the data would be skewed.

I believed that each color of M&Ms would not occur in equal amounts, so my alternate hypothesis was:

Ha: μBlue = μGreen = μOrange = μRed = μYellow = μBrown

My null hypothesis was:

H0: μBlue ≠ μGreen ≠ μOrange ≠ μRed ≠ μYellow ≠ μBrown

I set my alpha level at α=0.05.

Results

For my data analysis, I used Excel to calculate a Chi-Square, assessing a goodness of fit between my observed data and theoretically expected data. I also calculated descriptive statistics using Excel and used that to perform a confidence interval analysis which I plotted using a high-low-close graph. The results of my Chi- Square Expected Data are in Table 1.

Table 1

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Expected Data** | **Blue** | **Green** | **Orange** | **Red** | **Yellow** | **Brown** |
| 1 | 3.795731707 | 1.966463415 | 3.521341463 | 1.875 | 1.692073171 | 2.149390244 |
| 2 | 4.048780488 | 2.097560976 | 3.756097561 | 2 | 1.804878049 | 2.292682927 |
| 3 | 4.048780488 | 2.097560976 | 3.756097561 | 2 | 1.804878049 | 2.292682927 |
| 4 | 4.048780488 | 2.097560976 | 3.756097561 | 2 | 1.804878049 | 2.292682927 |
| 5 | 3.542682927 | 1.835365854 | 3.286585366 | 1.75 | 1.579268293 | 2.006097561 |
| 6 | 4.048780488 | 2.097560976 | 3.756097561 | 2 | 1.804878049 | 2.292682927 |
| 7 | 4.048780488 | 2.097560976 | 3.756097561 | 2 | 1.804878049 | 2.292682927 |
| 8 | 3.795731707 | 1.966463415 | 3.521341463 | 1.875 | 1.692073171 | 2.149390244 |
| 9 | 4.048780488 | 2.097560976 | 3.756097561 | 2 | 1.804878049 | 2.292682927 |
| 10 | 4.048780488 | 2.097560976 | 3.756097561 | 2 | 1.804878049 | 2.292682927 |
| 11 | 3.795731707 | 1.966463415 | 3.521341463 | 1.875 | 1.692073171 | 2.149390244 |
| 12 | 4.048780488 | 2.097560976 | 3.756097561 | 2 | 1.804878049 | 2.292682927 |
| 13 | 4.048780488 | 2.097560976 | 3.756097561 | 2 | 1.804878049 | 2.292682927 |
| 14 | 4.048780488 | 2.097560976 | 3.756097561 | 2 | 1.804878049 | 2.292682927 |
| 15 | 3.795731707 | 1.966463415 | 3.521341463 | 1.875 | 1.692073171 | 2.149390244 |
| 16 | 3.795731707 | 1.966463415 | 3.521341463 | 1.875 | 1.692073171 | 2.149390244 |
| 17 | 4.048780488 | 2.097560976 | 3.756097561 | 2 | 1.804878049 | 2.292682927 |
| 18 | 4.048780488 | 2.097560976 | 3.756097561 | 2 | 1.804878049 | 2.292682927 |
| 19 | 4.048780488 | 2.097560976 | 3.756097561 | 2 | 1.804878049 | 2.292682927 |
| 20 | 3.795731707 | 1.966463415 | 3.521341463 | 1.875 | 1.692073171 | 2.149390244 |
| 21 | 4.048780488 | 2.097560976 | 3.756097561 | 2 | 1.804878049 | 2.292682927 |

The result of my descriptive data analysis is summarized in Table 2.

Table 2

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Descriptive Statistics |  |  |  |  |  |
|  | *Blue* | *Green* | *Orange* | *Red* | *Yellow* | *Brown* |
|  |  |  |  |  |  |  |
| Mean | 3.952380952 | 2.047619048 | 3.666666667 | 1.952380952 | 1.761904762 | 2.238095238 |
| Standard Error | 0.387151936 | 0.223353131 | 0.347325382 | 0.223353131 | 0.238095238 | 0.291994471 |
| Median | 3 | 2 | 4 | 2 | 1 | 2 |
| Mode | 3 | 1 | 5 | 2 | 1 | 2 |
| Standard Deviation | 1.774153051 | 1.023532631 | 1.591644852 | 1.023532631 | 1.091089451 | 1.338086765 |
| Sample Variance | 3.147619048 | 1.047619048 | 2.533333333 | 1.047619048 | 1.19047619 | 1.79047619 |
| Kurtosis | -0.737321055 | -1.542344498 | -0.412232104 | 0.74354067 | -0.160572632 | 0.910785125 |
| Skewness | 0.435381891 | 0.206387186 | 0.115735948 | 0.102842992 | 0.778169591 | 1.041185912 |
| Range | 6 | 3 | 6 | 4 | 4 | 5 |
| Minimum | 1 | 1 | 1 | 0 | 0 | 0 |
| Maximum | 7 | 4 | 7 | 4 | 4 | 5 |
| Sum | 83 | 43 | 77 | 41 | 37 | 47 |
| Count | 21 | 21 | 21 | 21 | 21 | 21 |

The results of my confidence interval analysis are in Table 3

Table 3

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| M&Ms | **Blue** | **Green** | **Orange** | **Red** | **Yellow** | **Brown** |
| **Upper 95% CI** | 4.726684824 | 2.49432531 | 4.36131743 | 2.399087215 | 2.238095238 | 2.82208418 |
| **Lower 95% CI** | 3.178077081 | 1.600912785 | 2.972015904 | 1.50567469 | 1.285714286 | 1.654106297 |
| **Mean** | 3.952380952 | 2.047619048 | 3.666666667 | 1.952380952 | 1.761904762 | 2.238095238 |

The results of my Chi-Square and Chi-Square Probability is in Table 4.

Table 4

|  |  |
| --- | --- |
| **Chi-Sq=** | 79.45771387 |
| **ChiSq Prob =** | 0.064515204 |

The high-low-close plot is shown on the next page as Graph 1.

Graph 1



At first glance, looking at my high-low-close graph, which seems to show that blue and orange candies are more common, it would appear that my research would confirm what I had originally thought; that the colors of M&M candies in a fun sized bag do not occur with the same frequency. Though the P value of 0.0645, provided by my Chi-Square analysis, does not show the same conclusion. Being outside of the alpha level, set at 0.05, I am unable to reject my null hypothesis.

Conclusions

Given that the P value of the Chi-Square analysis is of near significance, I would suggest that this hypothesis be further researched by observing a larger data set. By collecting more data, I believe that the findings would show that the colors green and orange, indeed, occur more frequently.

I would also be interested in conducting a survey to find out whether or not people prefer a certain color M&M. By asking a random group of people to rank their favorite color M&M and comparing it to observed counts per bag of M&Ms it will be possible to compare the two data sets and determine if there is a correlation.

References

“vf-tropi.com” (2017).

SULLIVAN, M. (2014). Fundamentals of Statistics Fourth Edition. Pearson Education,

Inc.