2009

Skittles Chocolate Mix Color Distribution: A Chi-Square Experience

David R. Duncan
University of Northern Iowa

Bonnie H. Litwiller
University of Northern Iowa

Duncan, D. R., & Litwiller, B. H. (2009). Skittles Chocolate Mix Color Distribution: A Chi-Square Experience. Iowa Council of Teachers of Mathematics Journal, 36, 72-75.

SKITTLES CHOCOLATE MIX COLOR DISTRIBUTION: A CHI-SQUARE EXPERIENCE

David R. Duncan and Bonnie H. Litwiller

In teaching statistical processes, it is important that there be application to real-world settings and activities. When this is done, students are more likely to see the meaning of the steps being developed.

One such activity involves using the Chi-Square statistical test and its applications to counting Skittles Chocolate Mix candies. Many students are aware that these candies come in five different flavors: Brownie Batter (BB), Vanilla (V), Chocolate Caramel (CC), S'mores (S), and Chocolate Pudding (CP).

Let us first test the hypothesis that all colors are equally represented in this product. We will test this distribution hypothesis, called the Null Hypothesis, with four randomly selected 14-ounce bags of Skittles Chocolate Mix candies.

The following table reports the contents of these bags.

	NUMBERS			
FLAVOR	**Bag 1**	**Bag 2**	**Bag 3**	**Bag 4**
BB	89	65	83	65
V	72	117	75	114
CC	72	72	73	57
S	64	46	74	41
CP	79	73	73	92
TOTALS	**376**	**373**	**378**	**369**

Assuming generally equal numbers for the population of candies, color distributions should be as follows:

NUMBERS

FLAVOR	Bag 1	Bag 2	Bag 3	Bag 4
BB	75.2	74.6	75.6	73.8
V	75.2	74.6	75.6	73.8
CC	75.2	74.6	75.6	73.8
S	75.2	74.6	75.6	73.8
CP	75.2	74.6	75.6	73.8
TOTALS	**376**	**373**	**378**	**369**

To test the Null Hypothesis, we shall use the Chi-Square statistic. Let us construct Table 1 with column entries as follows for Bag 1:

O = The observed frequencies, the numbers of each color of Chocolate Mix candies actually present in our bag.

E = The expected frequencies (if the Null Hypothesis were true).

$(O-E)^2/E$ = A measure of the discrepancy between O and E.

Table 1:

FLAVOR	O	E	$(O-E)^2/E$
BB	89	75.2	2.53
V	72	75.2	0.14
CC	72	75.2	0.14
S	64	75.2	1.67
CP	79	75.2	0.19
TOTALS	**376**	**376**	**4.67**

In the last column (a measure of discrepancy), a small number indicates that O and E are relatively close together, as is the case for V. A larger number indicates that O and E are relatively far apart, as is the case for BB.

The sum of this discrepancy column, 4.67, is called the Computed Chi-Square Statistic (CCSS). A determination must be made as to whether the CCSS is large enough to cause us to reject the Null Hypothesis. To make this decision, a "referee" is needed. This referee is found in the Table Chi-Square Statistic (TCSS).

To read a Chi-Square table, the degrees of freedom must first be determined; that is, one less than the number of categories (colors). In our case, the degrees of freedom is 5-1 = 4. This means that if the total number of candies were known, and the number in each of four categories were known, the number in the fifth category could be calculated.

The significance level is the probability of rejecting a Null Hypothesis which is in fact true. This could occur because the sample is not representative of the population. From a Chi-Square table, we find:

SIGNIFICANCE LEVEL	TCSS
10%	7.78
5%	9.49
1%	13.28

The decision mechanism for the Null Hypothesis is:

- If CCSS > TCSS, then CCSS is large in the "judgment of the referee." If this is true, *reject* the Null Hypothesis.
- If CCSS < TCSS, then CCSS is small in the "judgment of the referee." If this is true, *accept* the Null Hypothesis.

Our CCSS of 4.67 is smaller than the TCSS's for any of the SL's. In other words, there is insufficient evidence to reject the Null Hypothesis for all three significance levels. According to our evidence, the assumption of equal numbers is accepted for Bag 1.

Repeating this Chi-Square analysis for the same Null Hypothesis for Bags 2, 3, and 4 yields CCS's of, respectively, 36.42, 0.93, and 45.84. For Bag 3, we again accept The Null Hypothesis of equal numbers in the product as a whole. However, for Bags 2 and 4 the CCSS's are larger than any of the TCSS's, leading us to reject the Null Hypothesis and to conclude instead that the varieties are unequally represented in the product as a whole.

What would happen if the contents of the four bags were combined? The following chart results:

FLAVOR	O	E	$\frac{(O-E)^2}{E}$
BB	302	299.2	0.03
V	378	299.2	20.75
CC	274	299.2	2.12
S	225	299.2	18.40
CP	317	299.2	1.06
TOTALS	**1496**	**1496**	**42.36**

Since this CCSS of 42.36 is much larger than any of the TCSS's, we confidently reject the hypothesis of equal population numbers. Bags 1 and 3 might have suggested to the contrary, but the pooled results decisively reject the Null Hypothesis.

The reader and his/her students are encouraged to investigate other distributions using this Chi-Square process.

David R. Duncan is a Professor of Mathematics at the University of Northern Iowa in Cedar Falls, Iowa. david.duncan@uni.edu

Bonnie H. Litwiller is a Professor of Mathematics at the University of Northern Iowa in Cedar Falls, Iowa. bonnie.litwiller@uni.edu