The Mysterious Case of the Blue M&M's®

Ronald D. Fricker, Jr.

Blue M&M's® are a new phenomenon on the candy scene, officially introduced in September of 1995. The color blue was chosen democratically when M&M/MARS, makers of the M&M's® Chocolate Candies, asked the public to vote for a new color. More than ten million people voted in this national election, and 54% of those chose blue over purple (32%), pink (10%), and "no change" (4%). Ten million people are equivalent to the combined populations of Kentucky, Tennessee, and West Virginia. Such a sizable voter turnout indicates that this is an area of research not to be taken lightly.

Blue replaced the "less flashy" tan in the M&M's® Plain Candies. As reported in *The Record* (7 July 1995, p. B01) Marlene Machut, an M&M/MARS company spokesperson, said, "tan was removed from the mix because, in our research, it was the least popular color. The company couldn't afford to add another color, but it could afford to replace one."

One recent Saturday, blissfully ignorant of these details, I was grading homework assignments with M&M's® at my side. Looking at the bowl of candy, I noticed that there seemed to be far fewer of the new blue variety than of the other colors. This led me to think about how estimating the proportions would make an interesting in-class experiment. Ever on the lookout for an excuse to avoid grading papers, however, I decided to conduct my own experiment right then and there.

Given that I had eaten most of the present data, I first headed out to the store. In the true spirit of the statistical profession, I applied the more-data-is-better principle and bought three pounds of candy as follows:

- from store A, one 16-ounce bag
- also from store A, one 10-ounce and one 6-ounce bag
- from store B, another 16-ounce bag

My rationale for this was to try and get three different one-pound samples that were not part of some common manufacturing lot (in fact, the lot numbers were different: 521DM12, 534DR3, 517GM14, and 525EM22, respectively).

Returning home I set about the task of counting the M&M's® with the goal of ensuring accurate counts while avoiding the temptation to munch on the data. My methodology was to count each bag individually, first separating the colors and then counting out piles of 10 M&M's® of each color. After reverifying the count in each pile, I made two counts of the piles, and was thus reasonably sure that I had accurate tallies by bag and by color. Table 1 lists the counts for each of the colors by pound and the aggregate total for all three pounds.

Table 1—The Data: Counts of M&M's® by Color for Three Different Pounds

Color	Pound #1	Pound #2	Pound #3	Total
Brown	177	146	132	455
Yellow	135	97	111	343
Red	79	124	115	318
Orange	41	69	42	152
Green	36	43	50	129
Blue	38	42	50	130
Total	506	521	500	1,527

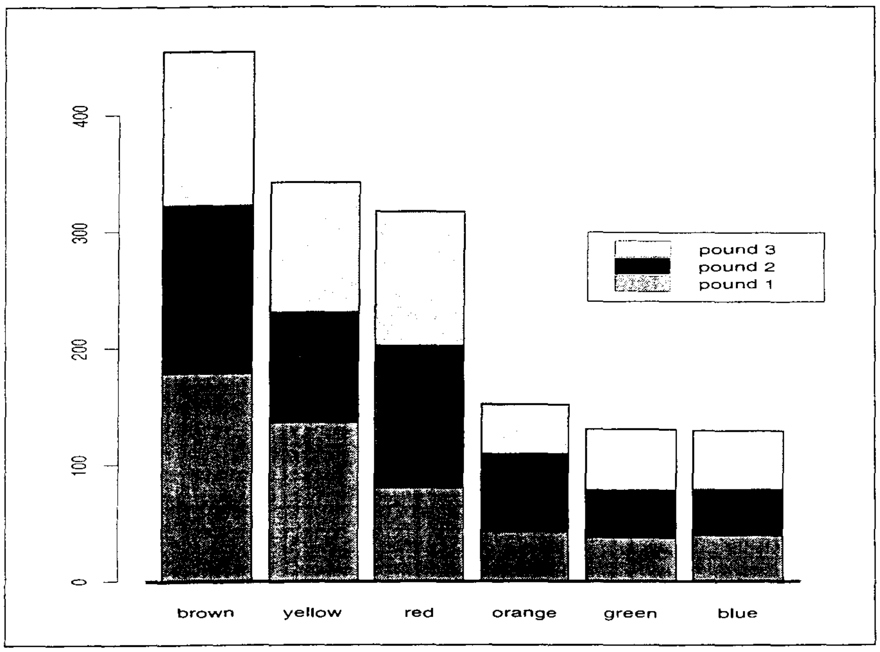

Figure 1. A simple bar chart of the data by color.

The first thing that struck me, when looking at the counts, was that they were quite consistent across samples (pounds) and that there seemed to be "classes" of colors. Most populous was the brown, followed by yellow and red, and then least populous were orange, green, and blue. A bar chart of the colors by pound in Fig. 1 confirms this insight.

Figure 2 shows simultaneous 95% confidence intervals for the proportion of each color. These confidence intervals were calculated via the methodology proposed by Quesenberry and Hurst (1964). Let $n_1,\ldots,n_k$ denote the aggregate count of each of k colors and N denote the total sample size; for example, if n_1 is the number of brown M&M's®, then for this data $n_1 = 455$ and $N = 1{,}527$. For π_i, the true proportion of color i, the simultaneous confidence intervals are calculated as

$$\pi_i = \frac{\chi^2 + 2n_i \pm \sqrt{\chi^2[\chi^2 + 4n_i(N - n_i)/N]}}{2(N + \chi^2)}$$

where $\chi^2 = \chi^2_{\alpha,K-1}$, for an interval of at least $1 - \alpha$ confidence (this procedure is conservative, with $1 - \alpha$ as the lower bound). For this example $\chi^2_{.05,5} = 11.0705$. Figure 2 shows that there are two distinct groups, with the existence of three groups a clear possibility.

Intrigued by the fact that the distribution of colors was not uniform and that they seemingly fell into at least two classes, I decided to do some background research. A NEXIS search on keywords "blue" and "M&M" quickly inundated me with information from the popular media. One particular article in *The Austin American-Statesman* (AAS), in response to an inquiry by a high school chemistry teacher, purported to provide the true distribution of colors in plain M&M's®. Table 2 lists the AAS percentages along with the percentages calculated from my data.

Clearly something was awry between my data and the alleged "true" distribution. Although most of the colors seemed to match up quite well, the brown and blue percentages were off. Indeed, in Fig. 2 the other four colors' observed proportions all fall within the calculated confidence intervals, but the brown and blue fall far outside.

With this discrepancy before me I decided to investigate the probability of observing 130 or fewer blue M&M's® under the hypothesis that the data was generated according to the AAS-alleged color distribution. Simplifying the colors into two groups, blue and not blue, the probability of X blues out of 1,527 M&M's® is binomial, and summing over $X = 1,\ldots,130$ gives the desired probability. Simplifying the calculations by invoking the normal approximation to the binomial gives

$$p(X \le 130) \approx$$
$$p(Z \le (130 - 1{,}527 x .2)/\sqrt{1{,}527 x .2 x .8})$$
$$= p(Z \le -11.2) = 2 \times 10^{-29}$$

Under the assumption that the M&M's® were part of a system that assigned the blue colors to the bags according to a binomial distribution with probability $p = .2$, observing 130 or fewer blue M&M's® out of a total of 1,527 M&M's® would virtually never occur.

Yet, this calculation really does not capture all that was observed with the data; not only are the blue counts approximately 10% lower, but the browns seem to be about 10% higher. The binomial calculation simply assumes that the blue underrepresentation was at the expense of any of the other colors, but Table 2 indicates that almost all of the missing blues fell into the brown category—an even rarer event.

To see this, apply a chi-squared goodness-of-fit test between the observed counts for the six colors and the expected counts according to the AAS distribution. The critical value for the chi-squared distribution with 5 degrees of freedom and $\alpha = .001$ is 20.5. Even at this small alpha level, the null hypothesis is strongly rejected with a calculated value of 182.9 and a resulting probability of 0 (more accurately, it was too small for S-Plus to calculate).

Thus, I could come to only one of two conclusions: Either I had uncovered a large corporate conspiracy designed to dupe an unsuspecting public out of blue M&M's®, or the newspaper was wrong. Hopeful of uncovering the former, I made a quick call to M&M/MARS and found the latter to be true. M&M/MARS consumer affairs stated that the true distribution contained 30% brown and 10% blue M&M's®. Feeling a bit foolish at first, I found out that M&M/MARS gets these sorts of inquiries from elementary and high school teachers quite often—so often, in fact, that they have created an entire brochure dedicated to discussing the distribution of colors in M&M's® candy!

Table 3 gives the distribution of colors in the various varieties of M&M's®. If we consider the M&M/MARS-reported distribution for the plain candies, a chi-squared goodness-of-fit test result of 12.2 fails to reject the M&M/MARS distribution for $\alpha = .01$. Now, although I could have rejected the null hypothesis with $\alpha = .05$, the "standard" hypothesis-testing value, care must be taken not to specify a test with too much power. What is important here is to note that the M&M/MARS distribution definitely fits the observed data better than the distribution reported by the AAS, where the AAS hypothesis would have been rejected for any α value chosen. And, note that all of the proportions reported by M&M/MARS fall in the confidence intervals of Table 2. We can rationalize the observed differences from the distribution reported by M&M/MARS by recog-

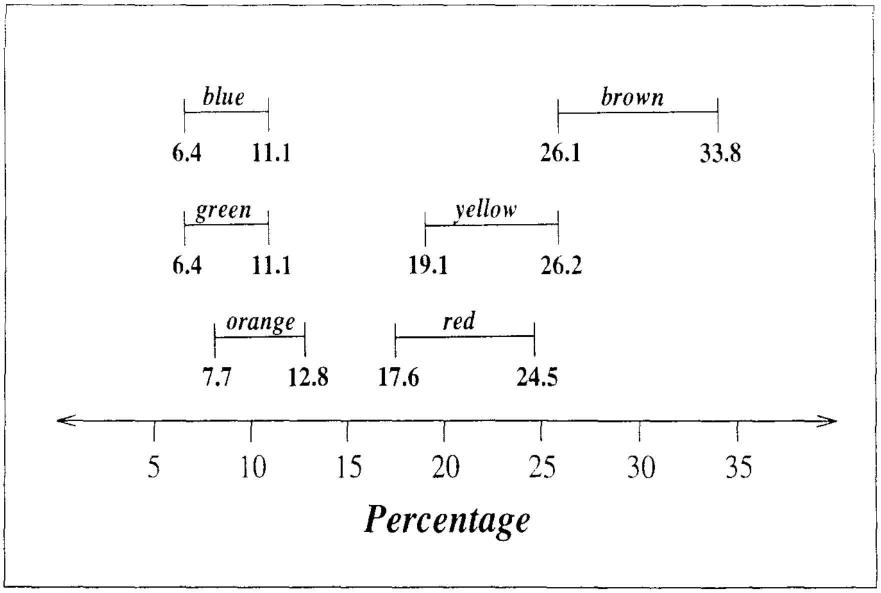

Figure 2. Approximate 95% confidence intervals for the true color proportions.

Table 2—The Percentage of Each Color Both Within Each Pound, for the Aggregated Pounds, and the Purported Distribution of the Colors as Printed in *The Austin American Statesman* (1 July 1995, p. D1)

Color	Sample data percentages: Pound #1	Pound #2	Pound #3	Aggregate percent	AAS percentages
Brown	35.0%	28.0%	26.4%	29.8%	20%
Yellow	26.7%	18.6%	22.2%	22.5%	20%
Red	15.6%	23.8%	23.0%	20.8%	20%
Orange	8.1%	13.2%	8.4%	10.0%	10%
Green	7.1%	8.3%	10.0%	8.4%	10%
Blue	7.5%	8.1%	10.0%	8.5%	20%
Total	100.0%	100.0%	100.0%	100.0%	100%

Table 3—The Percentage of Each Color for Various M&M Candies, as Reported in the Brochure "Colors" Distributed by M&M/MARS Consumer Affairs. Note That All the Easter Colors Are Pastels and so Are Actually Different from the Regular Colors

Variety	Brown	Yellow	Red	Orange	Green	Blue	Pink	White	Purple
Plain	30%	20%	20%	10%	10%	10%			
Peanut	20%	20%	20%	10%	10%	20%			
Almond	20%	20%	20%		20%	20%			
Peanut butter	20%	20%	20%		20%	20%			
Easter		20%			20%	20%	20%		20%
Valentine's			40%				40%	20%	
Christmas			50%		50%				

M&M's® as a Teaching Device

For instructors interested in using M&M's® as a teaching device, there are a wealth of possibilities, including (1) simple tests to illustrate basic concepts of sampling and variation, (2) more complex questions involving distribution estimation, (3) hypothesis testing, (4) acceptance sampling problems, and so forth. Resources in the literature that describe classroom activities in detail include Johnson (1993), Landwehr, Swift, and Watkins (1987), and the ASA Guidelines for Teaching Statistics in K-12.

The M&M/MARS "Colors" brochure provides additional details about how the colors are chosen, how they are applied to the candy, and the ratios of colors in each variety of the candy. Missing from the brochure is a description of the Halloween candy, for which readers will have to do their own analysis. The Colors brochure, as well as a one-sheet guide entitled "Suggested Mathematical Tasks Using 'M&M's®' Chocolate Candies and SKITTLES® Bite Size Candies" and the paper "Counting 'Taste Selections'," by Mitchell H. Taibleson, can be requested by writing to M&M/MARS Consumer Affairs, High Street, Hackettstown, NJ 07840, USA, or by calling (908) 852-1000.

nizing that true randomness must be at least slightly violated in the candy manufacturing and bagging process.

Table 3 shows that the distribution of colors for the peanut candy is different from the plain, and the common distribution for the peanut butter and almond candies is again different from both the plain and peanut distributions. This creates many new problem possibilities that I'll leave to interested readers with higher metabolic rates than mine. (And for those who find the standard colors too mundane, M&M/MARS has special colors for their Valentine, Christmas, and Easter confections.)

Even statisticians trained to be wary of the media can be caught off guard. *CHANCE News*, an electronic publication (http://www.geom.umn.edu/docs/education/chance/chance_news/news.html) dedicated to providing information on popular media stories containing interesting statistical issues (not to be confused with this publication), picked up on the AAS story and relayed it as fact. Forced to issue a correction, *CHANCE News* stated, "Well, even we put too much trust in newspapers!" (*CHANCE News*, 4.13).

Several insights resulted from this study. The first is that you can have your data and eat it too. Contrary to the usual maxim, though, too much data *can* be bad—either for your waistline, your complexion, or both, as well as for detecting unimportant differences. Second, skepticism is healthy when reading facts and statistics in the popular media. A cursory review of news reports about this simple topic uncovered numerous factual errors. Finally, and perhaps most importantly, the application of some simple statistical tools can help separate fact from fiction in our everyday lives, which leads me to wonder: Did the chemistry teacher ever figure out that *The Austin American-Statesman* was wrong?

References and Further Reading

Johnson, R. W. (1993), "Testing Colour Proportions of 'M&M's'®," *Teaching Statistics*, 15, 2–4.

Landwehr, J. M., Swift, J., and Watkins, A. E. (1987), *Exploring Surveys and Information from Samples* (Quantitative Literacy Series, 2 vols.), Palo Alto, CA: Dale Seymour Publications.

Quesenberry, C. P., and Hurst, D. C. (1964),"Large Sample Simultaneous Confidence Intervals for Multinomial Proportions," *Technometrics*, 6, 191–195.