Testing Colour Proportions of M&Ms

KEYWORDS:
Teaching; Simulation; Pearson's χ^2 statistic; Minitab; plain M&M's

Roger W. Johnson
Carleton College, Northfield,Minnesota U.S.A

Summary
We test the claimed colour proportions of candies using Pearson's χ^2 statistic. The null distribution of this statistic is examined through a Minitab simulation.

◆INTRODUCTION ◆

M&M'S® PLAIN CHOCOLATE candies, produced by Mars Incorporated, come in six different colours. According to a Mars (1991) consumer affairs representative "the color ratio. ..is 30% brown, 20% yellow, 20% red, 10% orange, 10% green and 10% tan [and] each large production batch is blended precisely to those ratios and mixed thoroughly." This claim may be tested in the classroom, of course, by distributing small sample bags of M&M's candies to students and having them count the number of each colour. The counts of the students can then be pooled and compared to the number expected with the claim being true. We use a goodness of fit statistic (Pearson's χ^2; see Pearson (1900) or Pearson (1956)) as a measure of how far the observed counts of the class differ from what is expected. To test the claim of Mars we estimate, using Minitab, the chance of seeing a goodness of fit statistic as extreme as the one we did assuming the claim is true. If this chance is small, we will have evidence to reject the claim of Mars. By using Minitab to estimate this chance an instructor dispenses with the need of explaining the use of a statistical table, namely the χ^2, to students.

◆DATA COLLECTION ◆

After discussing the claim of Mars above and eliciting response to the question how one might test this claim, I distribute small, "fun size", bags of M&M's candies to students in my class. While they are counting the number of each colour I discuss the need for carefully collecting data and how any conclusion we make assumes that the data have been accurately tabulated. Subtotals are obtained by the students at the end of each row, and I total these subtotals on the board in front of the class. The data in Table I, displayed along with the expected counts if Mars claim is true, are the result of such a classroom exercise:

Table 1.
Observed and Expected Counts

	brown	yellow	red	orange	green	tan	total
O	84	79	75	49	36	47	370
E	111	74	74	37	37	37	370

If Mars' claim is true, we would expect little deviation between the observed and expected counts. At this point one can introduce

$G = \Sigma(\text{Observed} - \text{Expected})^2/\text{Expected}$

(see Lehmann, 1986, pp 477-480, or Rice, 1988, pp 280-283) as a traditionally used measure of the goodness of fit between the observed and expected values. In our case

$$G = \frac{(84-111)^2}{111} + \frac{(79-74)^2}{74} + \frac{(75-74)^2}{74} + \frac{(49-37)^2}{37} + \frac{(36-37)^2}{37} + \frac{(47-37)^2}{37} = 13.54$$

Noting the general form of the numerator of the summand, we see that this statistic G will be large if any colour count is far from what is expected, and will be small if there is close agreement to observed and expected counts.

◆ SIMULATION ◆

Is the value G = 13.54 unusual if the claim of Mars is true? We can test this by assuming their claim is true and then computing the value of G for many different samples of size 370 from the claimed colour distribution. Using Minitab, for instance, one can do this by creating the text file 'mnms.mac' which follows (use the Minitab

STORE command or any word-processing package):

```
noecho
random k20 c3;
discrete c1 c2.
unstack c3 c4-c9;
subscripts c3.
let kl=n(c4)
let k2=n(c5)
let k3=n(c6)
let k4=n(c7)
let kS=n(c8)
let k6=n(c9)
set c 10
k1 k2 k3 k4 k5 k6
let c11=c2*k20
let c12=(c10-c11)*(c10-c11)/c11
let k10=sum(c12)
name c21 'brown' c22 'yellow' c23 'red'
name c24 'orange' c25 'green' c26 'tan' c27 'G'
stack k1 c2l c21
stack k2 c22 c22
stack k3 c23 c23
stack k4 c24 c24
stack k5 c25 c25
stack k6 c26 c26
stack k10 c27 c27
end
```

Once this has been done, the desired simulation can be performed with the following Minitab commands:

```
MTB >set c1
DATA> 1:6
DATA> end
MTB > set c2
DATA> 0.3 0.2 0.2 0.1 0.1 0.1
DATA> end
MTB > note: Set the number of M&M's to
sample
MTB > let k20 = 370
MTB > erase c21 -c27
MTB > note: Take, say, 100 samples of size 370
MTB > execute 'mnms.mac' 100
MTB > note: Display results
MTB > print c21-c27
MTB > sort c27 c28
MTB > print c28
```

[Comment: naming the columns inside the macro mnms.mac slows execution somewhat but reduces the typing required by the student if the instructor provides the macro.]

The (edited) output of the above 'print c21-c27' command, in one particular case, is given in Table 2.

Table 2. Simulation Colour Counts

Row	brown	yellow	red	orange	green	tan	G
1	114	70	75	29	45	37	3.7703
2	105	73	90	32	37	33	4.9054
3	119	73	72	38	36	32	1.3739
4	112	66	73	39	38	42	1.6982
5	119	74	71	32	33	41	2.2387
6	97	87	74	35	37	40	4.4009
7	119	57	84	39	37	34	6.1847
8	112	63	79	37	37	42	2.6577
9	105	80	72	51	24	38	10.7568
10	115	79	60	34	40	42	4.2928
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
97	115	74	75	36	31	39	1.2658
98	128	62	85	32	35	28	9.1577
99	110	75	70	35	39	41	0.8874
100	115	79	71	36	27	42	4.0090

From Table 2 it is seen that in the first computer sample of 370 M&M's there were 114 brown, 70 yellow 37 tan. The 'print c28' command prints out the 100 values of G associated with the 100 computer samples of size 370. These values are listed in Table 3.

Table 3. Simulation Goodness of Fit Statistics

0.7793	0.8874	1.0631	1.1216	1.1441
1.2658	1.3739	1.4685	1.6486	1.6982
1.8604	1.8739	1.9324	2.0766	2.2387
2.2387	2.2568	2.3468	2.4144	2.4550
2.5090	2.5270	2.5631	2.5811	2.6171
2.6171	2.6577	2.6892	2.7252	2.8108
2.8198	2.8333	2.9189	2.9820	3.1171
3.2117	3.2117	3.2703	3.4279	3.4459
3.4685	3.5000	3.5360	3.6441	3.6982
3.7703	3.8739	4.0090	4.0225	4.1171
4.1441	4.2928	4.4009	4.4144	4.4730
4.6171	4.6892	4.8874	4.9009	4.9054
5.0000	5.0360	5.1171	5.1216	5.1351
5.6036	5.6622	5.6622	5.9550	6.0225
6.1847	6.4865	6.5090	6.7297	6.8649
7.1036	7.3739	7.4820	7.5946	7.8604
7.9009	7.9550	8.0811	8.1892	8.2252
8.4144	8.6577	9.1577	9.1757	9.3739
9.5901	10.0541	10.2252	10.4009	10.6216
10.7568	10.7568	11.9009	12.1486	15.1622

This is a typical collection of values of G if the claim of Mars is true. In our classroom exercise, we saw G=13.54. Since only one value out of the 100, namely 15.1622, was bigger than that observed in class, we conclude that if the claim of Mars is correct, only about 1 time out of 100 will we see a value of 0 as big as 13.54. This gives us strong evidence to reject the claim of Mars. Each time I have conducted this classroom exercise, in fact, I have been led to the conclusion that the claim of Mars is incorrect. What should be noted here, however, is that the above analysis assumes that we have taken a simple random sample of M&M's candies. Consequently, if the colours are not as thoroughly mixed as claimed, we could easily reject the claim concerning the proportions of colours when, in fact, it is true.

References

Lehmann, E. (1986). *Testing Statistical Hypotheses, Second Edition.* John Wiley & Sons, New York.

Mars, Inc. (1991). Personal Communication with Brad Wagner, Consumer Affairs. High Street, Hackettstown, New Jersey 07840.

Pearson, K. (1900). On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from a random sampling. *The London, Edinburgh and Dublin Philosophical Magazine and Journal of Science,* fifth series, 50, 157-172.

Pearson, K. (1956). *Karl Pearson's Early Statistical Papers.* University Press, Cambridge.

Rice, J. (1988). *Mathematical Statistics and Data Analysis.* Wadsworth & Brooks/Cole, Belmont, CA.