A Sweet Demonstration of Statistical Hypothesis Testing

Jon Juneau/Karen R. Juneau/Eyler R. Coates, Jr
University of Southern Mississippi/Louisiana State University/University of Southern Mississippi

Abstract

More than eighty years ago, John Dewey warned that formal instruction becomes "remote and dead" when it is divorced from practical applications [3]. Many engineering topics are abstract so that it is difficult for instructors to effectively engage their students in these concepts. One method of clarifying abstract concepts is the use of simple simulation exercises. This paper describes a classroom demonstration of statistical hypothesis testing that tests a hypothesis about the distribution of the different colored candies in the "Fun Size" packages of M&M candies.

Introduction

Effective education builds upon the experiences of the individual. Students learn best when they can apply what they have learned. The application of knowledge to the solving of specific problems was an important factor in the founding of engineering and technology programs in this century [2]. More than eighty years ago, John Dewey warned that formal instruction becomes "remote and dead" when it is divorced from practical applications [3]. It is ironic that despite a considerable interest in assessing and recognizing individual learning styles that a common complaint about engineering education continues to be that the courses are often too abstract. Graduating students describe entering the work force without any real understanding of the concepts that they have been exposed to [4]. Many engineering topics are abstract and it is difficult for instructors to effectively engage their students in these concepts simply because these concepts are often foreign to student experience.

One method of clarifying abstract concepts is the use of simple simulation exercises such as the card drop shop exercise described by Alloway [1]. The card drop shop exercise illustrated quality concepts and provided an alternative to the red bead and funnel experiment that was used by Deming. This paper uses a similar approach to illustrate statistical concepts without any specialized or expensive equipment. It introduces students to basic statistical concepts and effectively engages student interest in these concepts.

Learning the concepts of statistical hypothesis testing is often presented in an abstract way. The meanings of the concepts of null and alternative hypotheses, Type I and Type II errors, significance levels and the critical regions are not easily grasped. To help the students digest these concepts, a hypothesis about the distribution of colored candies in a package of M&M candies is tested.

This paper describes the classroom demonstration using the "Fun Size" packages of M&M candies to test a hypothesis about the color distribution. The chi-square statistic is reviewed, since it was the test used for testing the goodness-of-fit between the hypothesized color distribution and the observed color distribution. Then, the color counts were done by the class, and the hypothesis tested. Finally, some conclusions on the effectiveness of the demonstration are presented.

The Class Demonstration

The demonstration begins with describing the hypothesis to be tested. The instructor states a hypothesis that is claimed to be from previous experience. For this demonstration, the hypothesis is that of all the colors in a package of M&M candy, there are twice as many brown M&Ms as there are of any of the other colors and all the other colors are equally represented. For example, if there are seven M&Ms in a package, there will be two brown M&Ms and there will be one each of the other colors: red, yellow, green, blue and orange. Stating the null hypothesis mathematically, if 2p is the proportion of brown M&Ms, then the proportion of red M&Ms is p, the proportion of yellow M&Ms is p, etc. The alternative hypothesis is then decided to be that there is some other distribution for the colors in a package of M&Ms.

After the null and alternative hypotheses are stated, the concepts of the Type I and Type II errors can be described. For example, the hypothesis is the current belief that there is twice as many brown M&Ms than the other colors. If after doing the statistical test, the hypothesis is rejected, then there is a chance that a mistake is being made. This is a Type I error, where the hypothesis is really true but the data gathered indicates that it is not. The probability that this occurs can be controlled with the statistical test by determining the significance level or alpha. The other type of error, a Type II error, occurs if the statistics indicates that we should not reject the null hypothesis when the alternative hypothesis is true. In this case, the error would be continuing to believe that there are twice as many brown M&Ms as there are the other colors when in reality, there is some other distribution for the colors in a package of M&Ms.

The next part of the presentation is a description of the statistical test to test the distribution of colors in a package of M&Ms.

The Statistical Test

Most often tests of hypothesis are concerned about single population parameters such as μ, σ^2 and p. However, there are tests that also determine if a population has a specified theoretical distribution. This class of tests is called "goodness of fit" tests because they are based on how good a fit one can have between the frequency of occurrence of observations in an observed sample and the expected frequencies obtained from the hypothesized distribution.

The most often used goodness-of-fit test is based on the chi-square distribution. If we create a histogram from our example, then a well-known theorem states that

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i} \qquad (1)$$

where $\chi^2$is a value of the random variable $\mathbf{X}^2$ whose sampling distribution is approximated very closely by the chi-square distribution. There are k cells in the histogram. The symbols o_i and e_i represent the observed and expected frequencies, respectively, for the i^{th} cell [5]. It should be noted that this test statistic should not be used unless the expected frequency of each histogram cell is at least five.

Equation 1 indicates that if the observed frequencies are close to the expected frequencies, then the chi-square random variable will be small, indicating a good fit. Since large values indicate poor fits, which lead to rejection of the hypothesis, then the test is a one-tailed test where the right tail is the area of rejection of the hypothesis.

Before one can use this statistic for hypothesis testing, the degrees of freedom must be determined. Again, a theorem states that, "The number of degrees of freedom in a chi-square goodness of fit test is equal to the number of cells minus the number of quantities obtained from the observed data that are used in the calculations of the expected frequencies" [5]. For example, consider the tossing of a die. If the die was tossed 120 times then the expected frequencies of the outcomes are given in Table 1. Note that each of the expected frequencies is greater than five.

Outcome	1	2	3	4	5	6
Frequency	20	20	20	20	20	20

Table 1: Expected Frequencies of 120 Tosses of a Die.

Note that all that was required from the observed data to calculate the expected frequencies was the total number of tosses. Thus, only one quantity (the total number of tosses) from the observed data was used. Therefore, the degrees of freedom was the number of cells (6) minus the number of quantities obtained from the observed data that are used in the calculations of the expected frequencies (1), or 5 degrees of freedom.

For the problem where the number of M&Ms is hypothesized to be twice as many browns as the other colors, the expected frequencies would depend only upon the total number of observations of the observed data. Thus, the degrees of freedom would be the number of cells minus 1. Since there are six different colors of M&M candies, the degrees of freedom for testing the null hypothesis will be 5.

Finally, once the level of significance, or alpha, is chosen, then many tables of chi-square critical values exist where one can determine whether the chi-square variable obtained leads one to accept or reject the hypothesis. For the demonstration, the value of alpha was chosen to be 5%. From a table of the chi-square distribution, the value calculated from equation 1 will have to be less than 11.1 to not reject the null hypothesis. In other words, if the value calculated is greater than 11.1, then there is less than a 5% chance that there are twice as many brown M&Ms as there are the other colors.

The Results of the Hypothesis Test

The next step in the demonstration is to pass out the "Fun Size" packages of M&M candies to the students. The students then open the packages and count the number of different colored M&Ms in their packages.

Table 2 shows the numbers for each color of M&M that the students had in their packages. The last column and row show the totals. For example, all 12 packages contained a total of 297 M&M candies.

From the observed frequencies of the different colors of the M&Ms and the expectation that twice as many M&Ms are brown, the chi-square statistic was calculated. With 297 M&Ms, the hypothesis would require that about 85 (84.86) should be brown, and there should be about 43 (42.43) of each of the other five colors. Using equation 1, the calculated chi-square value from this demonstration was 102.93 which is larger than the critical value of 11.1. Therefore, the null hypothesis that there are twice as many brown M&Ms as there are the other colors, had to be rejected.

Conclusions

The first conclusion that can be drawn from this demonstration is that there appear to not be twice as many brown M&Ms as there are of the other colors. Therefore, the null hypothesis must be rejected. This result provided the opportunity to describe how the scientific method generates knowledge by stating a hypothesis, testing the hypothesis, and when the results are unlikely, rejecting the hypothesis and develop a new theory. For example, the null hypothesis could be changed to browns and yellows have twice as many as the other colors the next time this demonstration is run.

Feedback from the students was collected during the following class period. One student indicated that the use the M&Ms "was a good hands-on example of how to do hypothesis testing." The use of the M&Ms in teaching hypothesis testing was an enjoyable way to demonstrate the abstract concepts of statistical hypothesis testing.

This type of study, of the M&M colors, was also used in a different class to illustrate the use of control charts. In this class, the blue M&Ms were considered as "non-conforming" and then the control limits for a p Chart were calculated. The results indicated that the proportion of blue M&Ms in the "Fun Size" packages of M&Ms appear to be in statistical control.

Brown	Red	Yellow	Green	Blue	Orange	Total
8	3	7	2	0	2	22
7	5	7	1	2	2	24
9	2	6	3	2	4	26
10	4	10	1	1	0	26
11	2	4	1	3	4	25
14	3	6	1	0	1	25
7	2	6	6	1	1	23
9	4	7	3	2	2	27
8	4	8	1	2	1	24
8	7	8	0	1	2	26
11	1	12	1	0	0	25
7	4	10	1	1	1	24
109	41	91	21	15	20	297

Table 2: The Class' M&M count

Acknowledgments

The authors are grateful for the comments of the reviewers which helped to improve the paper.

References

1. Alloway, J. A., (1994), "The card drop shop," *Quality Progress*, **2** (7).
2. Bennett, C. A., 1926, *History of manual and industrial education, 1870 to 1917*, Chas. A. Bennett Co., Peoria, IL.
3. Dewey, J., (1916), *Democracy and Education*, The Macmillan Company, New York.
4. Maul, G. P., (1994), "Reforming engineering education," *Industrial Engineering*, **26** (10) 53-67.
5. Walpole, R. E., and Myers, R. H., (1972), *Probability and Statistics for Engineers and Scientists*, The Macmillan Company, New York.

JON JUNEAU

Jon Juneau is currently a Visiting Assistant Professor of Industrial Engineering Technology at the University of Southern Mississippi in Hattiesburg, Mississippi. He received his B.S. and M.S. degrees in Nuclear Engineering from Texas A&M University in 1978 and 1980, respectively. He received his M.S. in Mathematics from Texas A&M University in 1981 and his Ph.D. in Engineering Science from Louisiana State University in 1996. He is a licensed P.E. and has thirteen years experience in nuclear fuel management for electric utilities. His research interests include quality improvement, particularly related to higher education, and the development of mathematical models and methods for improving industrial productivity.

EYLER ROBERT COATES

Eyler Robert Coates graduated from Louisiana State University in 1979 with a Bachelor of Science degree in Industrial Engineering. He has worked at Cincinnati Milacron, Hatteras Yachts and Davis Yachts as an industrial engineer. He obtained a Master of Engineering Science degree from Louisiana State University in 1996. He has completed his doctoral coursework at LSU and plans to complete his dissertation in 1998. Eyler Coates is currently an assistant professor at the University of Southern Mississippi in Hattiesburg.

KAREN R. JUNEAU

Karen R. Juneau is currently an Instructor of Vocational Education at Louisiana State University in Baton Rouge, Louisiana. She received her B.S. and M.S. degrees in Industrial Education from Texas A&M University in 1980 and 1981, respectively. She received her Ph.D. in Vocational Education from Louisiana State University in 1997. She has seven years experience as an engineering educator at Louisiana State University and has co-authored a workbook for engineering graphics. Her primary research interests include qualitative research methods and evaluation research methodology.