PROBABILITY SAMPLING AND INFERENTIAL STATISTICS: AN INTERACTIVE EXERCISE USING M&M'S*

CAROL J. AUSTER
Franklin and Marshall College

STUDENTS INTERESTED IN SOCIOLOGY soon discover that departmental requirements for an undergraduate major typically include a research methods course that incorporates statistics, a separate course on social statistics, or both. Despite the near inevitability of a quantitative component, students too often meet it with trepidation and a negativity that may interfere with their motivation, willingness, and ability to learn. These feelings include being intimidated by statistics and anxiety about their ability to understand course materials linked to quantitative data (Blalock 1987; Bridges et al. 1998; Potter 1995; Rushing and Winfield 1999; Schacht and Stewart 1990). One author acknowledged students' feelings by entitling his book *Statistics for People Who (Think They) Hate Statistics* (Salkind 2000) and another attempted to allay student fears with the title *Social Statistics Without Tears* (Johnson 1976). Faculty challenged by teaching statistics have suggested a variety of techniques designed to reduce students' apprehension and anxiety. Some suggest humor (Schacht and Stewart, 1990) and extended concrete examples (Rushing and Winfield 1999; Singleton 1989), while others recommend interactive gimmicks and class exercises (Markham 1991; Potter 1995; Schacht and Stewart 1992; Wybraniec and Wilmoth 1999) to alleviate students' anxiety, increase their confidence, and improve their comprehension of the material.

I focus here on teaching concepts associated with probability sampling theory and sampling distribution. Although students often come to a discussion of sampling design and methods with a long-standing familiarity with the words "random sample," they typically know little about probability sampling theory and the potential utility of a random sample or a sample drawn using other probability sampling methods. For example, most students can correctly calculate sampling error, but some remain mystified by the virtue of, or need for, such calculations. Such students wrongly assume that if a sample is selected using a probability sampling method, the size of the sample will not make any difference because the sample should be perfectly representative of the population. Although sociologists have suggested techniques for teaching students some of these specific concepts (Rushing and Winfield 1999; Singleton 1989; Wybraniec and Wilmoth 1999), the challenges associated with teaching sampling theory and concepts led me to develop this in-class exercise.

In the process of developing an in-class exercise to assure the participation of every student, I also had several learning objectives in mind. First, I wanted students to better understand the utility of probability samples, including the use of inferential statistics to estimate population parameters. Second, I expected that this hands-on exercise—particularly since we would have knowledge of the population parameter while working with the sample statistics—would lead to further understanding of concepts associated with samples and populations, including population parameter, sampling error, level of confidence, and confidence interval. Third, as a result of this exercise, I anticipated that students would be better able to interpret the results of opinion polls, understanding that despite the fact that

*Please address all correspondence to the author at the Department of Sociology, Franklin and Marshall College, Lancaster, PA 17604-3003; e-mail: c_auster@fandm.edu

Editor's note: The reviewers were, in alphabetical order, Suzanne Maurer, Steven P. Schacht, and Royce Singleton.

the actual population parameter is indeterminable, there are procedures available for its estimation. Finally, I hoped that a hands-on exercise that was illustrative, interesting, and fun would reduce negative feelings and anxiety about statistics and increase students' comprehension of relevant concepts as well as their confidence about, and comfort with, other statistical concepts. Given these objectives, although the exercise could be used in statistics classes, it seems more suited to research methods classes that tend to place more emphasis on the practical aspects of sample selection, including balancing the costs and benefits of selecting smaller or larger sample sizes.

PREPARING FOR THE EXERCISE

Before the class day in which we do the M&M's exercise, I lay the groundwork for students by discussing a variety of aspects of sampling and concepts associated with probability sampling theory as well as some of the inferential statistics associated with samples and populations (see, for example, Babbie 1998; Healey 1999; Levin and Fox 1999; Loether and McTavish 1993; Wybraniec and Wilmoth 1999). I begin by distinguishing between nonprobability and probability sampling methods and explain, at least theoretically, the basic principle of probability sampling: "a sample will be representative of the population from which it is selected if all members of the population have an equal chance of being selected in the sample" (Babbie 1998:200). I then illustrate how this basic principle is used by providing descriptions and concrete examples of some probability sampling methods including simple random sampling, systematic sampling, stratified sampling, and multi-stage cluster sampling. I offer several examples of how to calculate the sampling error using a binomial distribution, and explain that these calculations are only appropriate when a probability sampling method has been used. I explain that since we have only a sample from the population and not the whole population, we do not know the population parameter. Despite this, we can estimate the population parameter using the sample statistic and the sampling error. Using samples of different sizes in the calculations, I show how the sampling error decreases as the sample size increases.

A few concrete preparations also need to be made before the class meets for the M&M's exercise. First, for a class of about thirty students, two 16-ounce bags of regular M&M's (bags with brown, green, yellow, red, and blue M&M's) are needed. For each student in a class this size to have 20 to 25 M&M's, you will need about a bag and a half. With too few M&M's, the concepts are not as easily illustrated; yet, with too many M&M's, the exercise takes too long because students spend too much time counting them. Second, students should be asked to bring calculators to class. Third, if the instructor has access to a computer that has a software package with spreadsheet and charting capabilities, the calculations can be completed faster, easy comparisons can be made across the samples that will be drawn during the exercise, and instantaneous graphing will be available. If such capabilities are available, formulas can be entered into the spreadsheet either before the exercise to reduce the possibility for error and time needed or as the exercise progresses to further demonstrate the use of formulas in spreadsheets. If a computer setup is not available, most aspects of the exercise can still be carried out with calculators; however, a blackboard or overhead projector should be available to present the results.

CLASS PROCEDURES FOR THE M&M's EXERCISE

Step 1—Introduction and Initial Sampling

If you plan to use a computer spreadsheet for the calculation and presentation of some of the results, it helps to be in the classroom early to make sure the computer equipment is functioning properly. At the beginning of class, I welcome the students to an interesting, fun, and illustrative exercise involving none other than—M&M's! I explain that for the purposes of this exercise, all the M&M's in the room constitute the population. I then pass several cups of M&M's down the rows and ask each student to take a sample of 20 to 25 M&M's until they are all gone. I request that students not eat the "subjects" in their sample until the entire M&M's exercise is over! I ask them to count the number of yellow M&M's and the total number of M&M's in their possession and then to calculate the percent of yellow M&M's in their own sample by dividing the number of yellow M&M's by the total number of M&M's. They then record these numbers. The percent of M&M's that is not yellow is equal to 100 percent minus the percent of M&M's that is yellow. The binomial distribution for this dichotomous variable is represented by the percent of M&M's that is yellow and the percent of M&M's that is not yellow. This binomial distribution can be said to parallel, for example, the results of a poll that sampled respondents to find out what percent support each of two political candidates, or what percent support or oppose a local referendum.

Step 2—Seven Samples

Next, I ask the students to count off by sevens so that I have seven groups or samples from the M&M's population. I assign those seven groups (the 1s, the 2s, the 3s, etc.) to different areas of the room to count their M&M's and make the appropriate calculations. Specifically, the students in each of the seven groups are asked to pool their calculations; that is, to divide the number of yellow M&M's for the group by the total number of M&M's for the group to calculate the percent of yellow M&M's for the entire group. Although the results can be posted on a blackboard or on an overhead projector transparency, I find it most useful to enter the results in a spreadsheet so that they can be used for subsequent calculations (the results for these seven samples from the most recent exercise can be seen in Table 1). At this point, we also calculate the population parameter. In the most recent exercise, 22.9 percent of M&M's in our population of 673 were yellow. Using these results, we discuss the variation in the percent of yellow M&M's across the seven samples, and note the distance between the results from the samples and the actual population parameter. This reinforces the notion that although these are, in fact, probability samples, they are not perfectly representative of the population. Next, we calculate the sampling error (se=square root (p x q)/n) for each sample and spend some time looking at the variation in the sampling error. We note how the sampling error varies by both the sample size, ranging in Table 1 from 87 to 111, and the particular binomial distribution, which ranged from 18.4 percent/81.6 percent to 32.4 percent/67.6 percent. We then use the sampling error to calculate the confidence intervals at each of the three confidence levels, 68 percent (+-1se), 95 percent (+-2se), and 99 percent (+-3se). For example, at the 95 percent confidence level, the confidence interval for a sample with 18.4 percent yellow M&M's and a sampling error of 4.2 would be 18.4 percent +- (2 x 4.2%) which is equal to a confidence interval of 10.0 percent to 26.8 percent (see first row of Table 1). Theoretically, one can be 95 percent confident that the population parameter falls between 10.0 percent and 26.8 percent. We subsequently look at all of the confidence intervals for the first seven groups. The results of our recent class show that the population parameter actually fell within the confidence interval for only two of the seven samples at the 68 percent confidence level, and within the confidence interval for six out of seven samples and all seven samples at the 95 percent and 99 percent confidence levels, respectively (see Table 1). Occasionally, the confidence interval may seem one-tenth too high or too low because of rounding.

Step 3—Four Samples

At this point, I ask the class to count off by fours so that the sample groups are again mixed up and the number of M&M's in each sample is larger. The students in each of the four groups are again asked to pool their

Table 1. Sample Statistics and Data for Seven, Four, and Two Group Samples

Number of Samples	Sample Number	Number of M&M's n	Number of Yellow M&M's	Percent of Yellow M&M's p	Percent of Other M&M's q	Sampling Error se	Confidence Intervals: 68% Confidence Level -1se	+1se	95% Confidence Level -2se	+2se	99% Confidence Level -3se	+3se
Seven	1	87	16	18.4	81.6	4.2	14.2	22.6	10.0	26.8	5.8	31.0
	2	92	17	18.5	81.5	4.0	14.5	22.5	10.5	26.5	6.5	30.5
	3	83	15	18.1	81.9	4.2	13.9	22.3	9.7	26.5	5.5	30.7
	4	100	19	19.0	81.0	3.9	15.1	22.9	11.2	26.8	7.3	30.7
	5	98	22	22.4	77.6	4.2	18.2	26.6	14.0	30.8	9.8	35.0
	6	102	29	28.4	71.6	4.5	23.9	32.9	19.4	37.4	14.9	41.9
	7	111	36	32.4	67.6	4.4	28.0	36.8	23.6	41.2	19.2	45.6
Four	1	166	33	19.9	80.1	3.1	16.8	23.0	13.7	26.1	10.6	29.2
	2	187	43	23.0	77.0	3.1	19.9	26.1	16.8	29.2	13.7	32.3
	3	146	35	24.0	76.0	3.5	20.5	27.5	17.0	31.0	13.5	34.5
	4	174	43	24.7	75.3	3.3	21.4	28.0	18.1	31.3	14.8	34.6
Two	1	358	75	20.9	79.1	2.2	18.7	23.1	16.5	25.3	14.3	27.5
	2	315	79	25.1	74.9	2.4	22.7	27.5	20.3	29.9	17.9	32.3
Total Population		673	154	22.9								

calculations; that is, to divide the number of yellow M&M's by the total number of M&M's for the group to calculate the percent of yellow M&M's for the entire group. (The most recent results can be seen in the second section of Table 1). Using the results for these four samples, we discuss the variation in the percent of yellow M&M's across the four samples. With the results in Table 1, we note that the extremes seen in the previous smaller samples are not evident when the sample sizes are somewhat larger. We again note that some of the samples produce results quite close to the population parameter, but none produce the exact percent of yellow M&M's in the population. This again reinforces the need for sampling error statistics and the notion that even probability samples are not perfectly representative of the population. For the four samples, we then calculate the sampling error. This time we look not only at the variation in sampling error across these four groups, but also at how the sampling error for each of these four samples is less than that for any of the smaller samples produced when the class is divided into seven sample groups. We then use the sampling error to calculate the confidence intervals at each of the three confidence levels. The results in Table 1 show that the population parameter actually fell within the confidence intervals at all three confidence levels for each of the four samples.

Step 4—Two Samples

At this point, I ask the class to count off by twos so that the resulting two sample groups are again mixed up and the number of M&M's in each of these two samples is larger than in each of the previous four samples. Again, the students in each group are asked to pool their calculations. (See Table 1 for these results). We discuss the results of these two larger samples, noting that they seemed to be closing in on the population parameter. We note again that the extremes seen in the smaller samples, particularly those shown in the top section of Table 1, are not evident now that the sample sizes are the largest so far. We do find, however, that even these larger samples are not perfectly representative of the population. For these two larger samples, we calculate the sampling error. We note how the sampling error for these two samples is less than that of any previous samples. We again use the sampling error to calculate the confidence intervals for each of the three confidence levels. The results in Table 1 show that the population parameter fell within the confidence intervals at all three confidence levels for these last two samples.

Additional Considerations

To illustrate additional concepts associated with sampling distributions, particularly that of a normal curve, the following process can be implemented. Each student can be asked to report the percent of yellow M&M's in their own individual sample. With access to a computer with spreadsheet and charting capabilities, you can easily enter the results in a spreadsheet (such as Excel) with the percent of yellow M&M's (18%, 19%, 20%, etc.) listed in the first column and the number of students (1, 2, 3, etc.) who found that percent of yellow M&M's in their own sample in the corresponding second column. You can quickly plot the results with a chart function (such as "xy chart" in Excel), using a listing of the percent of yellow M&M's (18%, 19%, 20%) as the x-axis and the number of students who have that particular percent of yellow M&M's in their sample as the y-axis, or with the bar chart option, drawing a curved line over the bars in the chart (or "smooth lines" in Excel). If relying only on the use of a calculator, you can instead ask students to chart the percent of yellow M&M's in their samples on the board. As Wybraniec and Wilmoth (1999) suggest, students can place an 'X' on the line above the number that represents their sample results. "If another student has already placed an 'X' on that value, students must put an 'X' above the previous X" (Wybraniec and Wilmoth 1999:78). Regardless of which method is used to graph the data, the result should hopefully resemble a normal curve so that students can better understand sampling theory and distribution. A discussion of how the samples are distributed around the population parameter would augment what has already been illustrated by the rest of the exercise.

CONCLUSION

One of the advantages of this exercise is that it can be used with a small or large class. With a larger class, you can always provide fewer M&M's per person or start with more sample groups in order to assure the smaller sample sizes needed to produce larger sampling errors. With a smaller class, you can carry out the exercise as already suggested, or have each student draw more than one sample of their own, each of which would later be used as part of the larger samples. This would assure that there are plenty of samples and M&M's in the later stages of the exercise.

Although I have used the exercise to reinforce concepts associated with sampling and population, the exercise can be used as part of the initial presentation of, or introduction to, such concepts. In addition, to further illustrate these concepts, the M&M's exercise could be followed with an example that includes people's opinions rather than the color of candy. For example, analyzing the results of a recent *New York Times*/CBS news poll or a poll associated with a local, state, or national political election that includes the sampling error, would show students how they can use what they have just learned to enhance their understanding of those results. This could lead to further discussion of the relationship between sample size and sampling error, and the ways that researchers need to balance the costs in time or money or both with the level of sampling error they find acceptable. This is a good time to encourage discussion of insights garnered from this exercise in terms of the sampling constraints one might face when conducting research in an applied setting. Moreover, it is beneficial for students to be reminded of the differences between non-probability and probability samples and also, for example, between a simple random sample and a systematic sample.

There are also a variety of potential pitfalls that may be encountered when conducting this exercise. First, although few students make errors calculating the percentage of yellow M&M's, you may want to have students recheck their own calculations and those of their group for sampling error and confidence intervals to make sure the correct results are reported. Second, since these are real samples of M&M's, the particular distribution of yellow M&M's may be problematic. For example, the results for some of the smaller samples may more closely approximate the population parameter than the results for the larger samples. Or, as a result of the sample mean and the sampling error, an unusually high number of the smallest samples may include the population parameter at the 68 percent confidence level (+-1se) compared to some of the larger samples. Although this occurs infrequently, the nature of probability sampling means that such results are at least a possibility. If this happens, students could discuss why it occurred and the likelihood of it recurring. To explore this in even greater detail, you could redistribute the M&M's at any stage of the exercise, as long as the students have not eaten them. This brings me to an additional, less academic pitfall. Since hungry students may cause the sample sizes to dwindle as the exercise progresses, be careful about the time of day this exercise is scheduled! This is not a problem, however, if there is no need to resample or recalculate.

I have used this exercise twice. The second time, as part of a questionnaire about my research methods course, I asked two close-ended questions concerning the M&M's exercise: "How useful was the M&M's exercise in terms of understanding the representativeness of samples, particularly those of different sizes?" and "How interesting did you find the M&M's exercise?" Of the 27 students who responded, 74 percent indicated that the exercise was "very useful," 26 percent indicated that it was "somewhat useful," and none of the students indicated "not very useful." Although one student indicated the exercise was "not very interesting," the rest of the class thought that the exercise was either "very interesting" (52%) or "somewhat interesting" (44%). In informal conversations, students claim to have a much better understanding of samples and population and the reasons one needs a statistic such as sampling error. As a result of this exercise using M&M's, my students also seem less anxious and more comfortable with a variety of statistical concepts associated with samples and populations. In addition, at the end of the exercise, students satisfy their chocolate cravings when they consume their "subjects"—namely the M&M's!

REFERENCES

Babbie, Earl. 1998. *The Practice of Social Research*. Belmont, CA: Wadsworth.

Blalock, H. M., Jr. 1987. "Some General Goals in Teaching Statistics." *Teaching Sociology* 15:164-72.

Bridges, George S., Gerald M. Gillmore, Jana L. Pershing, and Kristin A. Bates. 1998. "Teaching Quantitative Research Methods: A Quasi-Experimental Analysis." *Teaching Sociology* 26: 14-28.

Healey, Joseph F. 1999. *Statistics: A Tool for Research*. Belmont, CA: Wadsworth.

Johnson, Allan G. 1976. *Social Statistics without Tears*. New York: McGraw-Hill.

Levin, Jack and James A. Fox. 1999. *Elementary Statistics in Social Research*. Reading, MA: Addison-Wesley.

Loether, Herman J. and Donald G. McTavish. 1993. *Descriptive and Inferential Statistics: An Introduction*. Boston, MA: Allyn and Bacon.

Markham, William T. 1991. "Research Methods in the Introductory Course: To Be or Not to Be?" *Teaching Sociology* 19: 464-71.

Potter, Alisa M. 1995. "Statistics for Sociologists: Teaching Techniques That Work." *Teaching Sociology* 23:259-63.

Rushing, Beth and Idee Winfield. 1999. "Learning About Sampling and Measurement by Doing Content Analysis of Personal Advertisements." *Teaching Sociology* 27:159-66.

Salkind, Neil J. 2000. *Statistics for People Who (Think They) Hate Statistics*. Thousand Oaks, CA: Sage.

Schacht, Steven and Brad Stewart. 1990. "What's Funny About Statistics? A Technique for Reducing Student Anxiety." *Teaching Sociology* 18:52-56.

_____. 1992. "Interactive/User-Friendly Gimmicks for Teaching Statistics." *Teaching Sociology* 20:329-33.

Singleton, Royce, Jr. 1989. "On Teaching Sampling: A Classroom Demonstration of Concepts, Principles, and Techniques." *Teaching Sociology* 17:351-55.

Wybraniec, John and Janet Wilmoth. 1999. "Teaching Students Inferential Statistics: A 'Tail' of Three Distributions." *Teaching Sociology* 27:74-80.

Carol J. Auster is a professor of sociology at Franklin and Marshall College. She recently completed *The Sociology of Work and Occupations: Syllabi and Other Instructional Materials* for the American Sociological Association's Teaching Resources Center. Her current research interests include the impact of gender on the career persistence and division of household labor among women and men in engineering.

Copyright of Teaching Sociology is the property of American Sociological Association. The copyright in an individual article may be maintained by the author in certain cases. Content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.