Bernoulli Trials: Do They Exist?

Robert L. Wardrop, Department of Statistics, University of Wisconsin-Madison

Key Words: Statistical education; Bernoulli trials; active learning; small student projects.

1 INTRODUCTION

My introductory statistics course emphasizes active learning and small student projects. For details on this approach, see Rossman [1], and Wardrop [2], [3], and [4].

This paper describes some of my experiences with student projects related to Bernoulli trials. In general, I believe that student projects that are *comparative studies* are more effective at stimulating student interest and creativity, and at teaching important statistical concepts, and that projects related to Bernoulli trials are less effective at achieving these results. Fortunately, however, one need not choose one type of project at the exclusion of another. Building mathematical models is an important area of statistics, but it is an area that, in my experience, undergraduates have difficulty understanding.

2 THE NEED FOR A MODEL

A finite population is a well-defined collection of subjects of interest. An infinite population is a mathematical model for the process that generates the outcomes of a sequence of trials. My students have relatively little trouble grasping the concept of a finite population, but they have substantial difficulty with the idea of an infinite population.

I know teachers who "solve" this difficulty by ignoring it. They proceed as follows. A popular way to conceptualize a finite population is as a collection of cards in a box, with one card per population member. These teachers simply tell the student that the box of cards can be used for trials too. For example, imagine a big box that represents all possible free throws attempted by Michael Jordan. If we observe Michael Jordan shooting free throws we can conceptualize the outcomes as a sample from the box.

This is not an unnatural way to proceed. If the goal of the course is for the student to respond properly to a given stimuli (that is, choose the proper formula and plug the appropriate number(s) into it with said stimuli including a careful statement of exactly which assumptions the student may use), this approach works fine. If, however, the course includes as a goal the student developing the ability to critically assess the application of a statistical method in a given situation, then it may prove useful to devote some class time to a more intellectually complete development of trials.

The following represents a typical application of statistical methodology for a finite population.

> A sample of 100 persons is selected from a population of size one million. Each person in the sample is asked the same question; 70 answer yes and 30 answer no.

On the assumption that the researcher selected a random sample from the population, one can compute a 95 percent confidence interval for p:

$$0.70 \pm 1.96\sqrt{\frac{0.70(0.30)}{100}} = 0.70 \pm 0.09 =$$
$$[0.61, 0.79].$$

The following is a typical application for trials.

> Sally shoots 100 free throws and obtains 70 successes.

On the assumption of Bernoulli trials, one can compute a 95 percent confidence interval for p and will obtain the answer given above for the random sample from the finite population.

An advanced graduate student at Wisconsin, who happened to be my teaching assistant, said to me,

> Why do you bother teaching your students about two types of populations when they both lead to the same formula?

Yes, indeed they do, but in my class, statistics is about using math to do science, not about doing math and pretending it is applicable by trotting out sanitized scientific examples. The focus in my course is on science; as science, polling a sample of people is quite different than shooting free throws.

Consider the different challenges the two researchers face. For a finite population, even in mathland one must deal with the difficulties inherent in actually selecting a random sample from the population, and then getting the sampled members to participate in the study. Even given a random sample, outside mathland the researcher must be concerned with any nonsampling biases that can invalidate the study, such as bias caused by the wording of a question. By contrast, the researcher studying trials has an easy life. Simply observe the outcomes of the trials and assume the process is Bernoulli trials! Easy, but not very intellectually honest and not necessarily valid. The researcher studying trials needs some tools to help decide whether the assumption of Bernoulli trials is reasonable.

Let us return to the two examples above with the same numbers, 70 successes in a sample of 100. It is my experience as a teacher that,

1. for finite populations:
   - students find generalization from sample to population natural, and
   - they understand the difference between p and $\hat{p}$.
2. but for infinite populations:
   - students find generalization from sample to process unnatural, and
   - they confuse p with $\hat{p}$; many believe they are the same number.

One reason students have difficulty understanding the need for Bernoulli trials is that typically Bernoulli trials are introduced in the context of the one sample problem. Two cases are considered. First, p is assumed to be known, which typically is uninteresting. Second, p is unknown and the student learns how to estimate p or test a hypothesis about its value. And here is the flaw in this approach. The student has not been convinced that knowing about p is *interesting*; the student has no reason to *care* about the probability of success for the underlying process. This notion of probability of success appears to be a mathematical convenience with no utility for solving interesting real problems.

I use the following two examples to show my students the value of studying Bernoulli trials.

1. Sally shoots 100 free throws and obtains 70 successes; the next day she shoots 100 free throws and obtains 75 successes. What should we conclude? Has Sally's ability to shoot free throws changed?
2. On Monday, Ralph attempts 100 six-foot golf putts and obtains 60 successes. He plans to attempt 75 six-foot golf putts on Tuesday. On the assumption that Ralph's putting ability has not changed, how many successes should we predict for Tuesday?

Note that the first of these is a two sample problem while the second considers a nonstandard topic in the study of one population—the problem of prediction. By this point in my course, the students have been exposed to probabilistic reasoning through hypothesis testing based on the randomization distribution and they can see the value of a mathematical model for answering these questions.

3 THE MIND EXPERIMENT

The three assumptions of Bernoulli trials are:

1. Each trial has two possible outcomes: success (S) or failure (F).
2. The probability of success, p, remains constant from trial to trial.
3. Trials are statistically independent.

The canonical example of Bernoulli trials is tossing a coin (fair or weighted) and I use it in my class. Clearly, each toss has two possible outcomes, heads or tails. Further, *clearly* the coin does not change during experimentation nor does it have a memory. With this argument I convince nearly all of the students in my class that tossing a coin is Bernoulli trials without ever producing any experimental results! This is an example of a *mind experiment*. Without any reality checking, we have drawn a conclusion about how a coin behaves. To be fair, some of my students may be drawing upon their past experiences with coins, but I do ask them simply to *think about* how a coin *should* behave.

Mind experiments certainly are useful and I advocate their use in teaching. The mistake, I believe, is in stopping at mind experiments. I urge my students to examine data critically to determine whether the assumption of Bernoulli trials is reasonable. Of course, to urge my students to be critical without providing guidance on how to examine data would be pointless. In fact, I have observed an interesting dichotomy among my students.

I say to my class, "Tonight I will go home and shoot 100 free throws. Are these Bernoulli trials?" Nearly all of my students who are math major answer yes, but a large majority of my other students answer no. The latter group has all kinds of reasons for answering no—I will improve from practice or perform worse because of fatigue, or I will experience hot and cold streaks. I conjecture that the math students answer yes simply because their math training does not devote much or any time to questioning assumptions. Mathematicians start with assumptions and deduce their logical consequences. After a brief class discussion I mention that none of my students has actually ever seen me shoot free throws; how dare they make pronouncements about me!

4 BEYOND THE MIND EXPERIMENT

I require my students to perform small data collection and analysis projects. Typically, I give my students several types of projects to choose from, with emphasis on comparative studies with a dichotomous or numerical response and either complete randomization or randomized pairs. Another option is to perform or observe a sequence of dichotomous trials and critically examine the assumption of Bernoulli trials. The Bernoulli trials projects are much less popular than the comparative studies, but over the years I have compiled a small collection of very good projects. In the remainder of this paper I will describe some of my favorite projects.

In her project paper, Kristen Joiner wrote,

> I chose to test Muffin's preference for balls. She has two balls we often throw for her to fetch. One is small and blue, the other is slightly larger, heavier and red. My father rolled both balls for her with one hand at the same time; I recorded which ball she chose to chase first. Arbitrarily, the blue ball was labeled a success.

Muffin participated in 96 trials and obtained a total of 57 successes (59 percent of 96). Are these Bernoulli trials? I do not know. (Teachers who cannot admit to limitations should not assign projects.)

I do not seek to give my students definitive, absolute answers. They do not exist in science. Instead I suggest two *simple* ways to examine the data.

First, one can compare the first and second halves of the study. In the current study, Muffin obtained 27 successes during the first half of the study and 30 successes during the second half, an increase in the success rate of six percentage points. At this point I am not interested in having my student formally assess the statistical significance of these data; instead I want them to think about them. Is a six percentage point improvement important or striking?

In addition, I have my students create the 2×2 table relating each trial to its immediate predecessor. For Kristen's study, 61 percent of Muffin's successes were followed by another success, and 56 percent of Muffin's failures were followed by a success.

From these data I conclude that there is little evidence against the assumption of Bernoulli trials. At this time I stress two ideas. First, this is my conclusion and the student is free to disagree. Second, I really cannot say that we have Bernoulli trials. I have simply failed to find much evidence *of a particular sort*. The value of p might have changed drastically during the course of the study; I compared only the first and second halves. In addition, there might have been all sorts of memory in the system; I simply examined one-step memory.

In his project paper, Eric Statz wrote,

> My dog Boone loves popcorn, so I thought if I throw pieces of popcorn for her to catch the assumptions of Bernoulli trials would be reasonable. While sitting at the dining room table I tossed one piece of popcorn every five seconds, approximately the same height and distance for her to catch.

Boone improved from 66 percent successes in the first 50 trials to 76 percent successes in the last 50 trials. More striking, 66 percent of Boone's successes were followed by a success, and 83 percent of Boone's failures were followed by a success. Did Boone feel bad after missing and subsequently try harder? Or did Eric subconsciously put more arc on tosses that followed a miss, making a success easier for Boone?

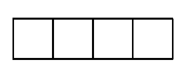
The descriptive comparisons I suggest can be made more formal by using hypothesis testing. If one does this for Eric's study the comparison of the halves of the study yields a one-sided P-value of 0.1891, and the investigation of memory gives a one-sided P-value of 0.0702.

For his class project Ben Relles obtained data from the 58 *Black Crowes* concerts that were performed between January 9, 1995, and September 4, 1995. A concert was labeled a success if it included the three songs *My morning song*, *Wiser time*, and *Remedy*, and was labeled a failure if at least one of these songs was not played. (Ben showed tremendous initiative in performing his project. He polled *Black Crowes* fans on the internet to determine the "people's choice" definition of a success.)

Ben's success rate increased from 38% during the first 29 concerts, to 52% during the last 29 concerts (one-sided P-value of 0.2143). Nineteen percent of successes were followed by a success, but 65% of failures were followed by a success (one-sided P-value of 0.0003).

Ben was not surprised by the memory in the process. The *Black Crowes* does not play one huge concert per city, instead it typically plays several concerts to smaller audiences at the same site. Before collecting data Ben conjectured that to avoid boring persons who attend two or more consecutive concerts, the band varies its selection of songs quite a lot. Thus, Ben conjectured there would be evidence of one-step memory in the data.

I performed my own project on dichotomous trials. I was looking for an example of Bernoulli trials to convince my non-math majors that they really do exist. I decided to study my favorite video game, Tetris. I defined a trial to be every time the program caused a shape to be dropped from the top of the screen. There are seven different shapes and I defined everyone's favorite, the rectangle,



to be a success, and any other shape to be a failure.

I had played Tetris for months and, even though I had never collected data, I was confident that I had Bernoulli trials. Imagine my surprise when I obtained the data displayed below.

Previous Shape	Current Shape S	F	Total
Success	2	251	253
Failure	249	1362	1611

Less than one percent of successes were followed by a success, but 15.5 percent of failures were followed by a success. There is most definitely one-step memory in Tetris!

Finally, Mieke Baseman, a student of the clarinet, wrote in her project report,

> One of the hardest things to learn [about playing the clarinet] is to consistently play in tune.

Mieke defined a trial to be playing a concert B$\flat$ into a tuner which classifies the note as perfectly in tune (a success), sharp (a failure), or flat (a failure).

Mieke showed considerable improvement over the course of her study, with 32 successes in the first 50 trials followed by 41 successes in the second fifty trials (the one-sided P-value is 0.0352). In addition, 81 percent of Mieke's successes were followed by a success, but only 58 percent of her failures were followed by a success (the one-sided P-value is 0.0216).

Mieke's data illustrate an important point. I am convinced that Mieke's trials were not Bernoulli trials, but I am not sure which assumption is violated. If p changes substantially from the first to second half of the study, the memory table will show a pattern. Conversely, a strong one-step memory might well result in a table that suggests p changed. This is not a fatal difficulty; even though I am not confident in naming which assumption is violated (or, if both are violated), I am comfortable in concluding that I should not treat the data as the outcomes of Bernoulli trials.

References

[1] Rossman, A.: Book Review of *Statistics: Learning in the Presence of Variation*, *The American Statistician*, May, 1995, Volume 49, Number 2, pages 237–8.

[2] Wardrop, R.: "A Radically Different Approach to Introductory Statistics," University of Wisconsin–Madison, Department of Statistics Technical Report No. 889, 1992.

[3] Wardrop, R.: "Guest Editorial: A New Approach to Introductory Statistics," *The Journal of Undergraduate Math and Its Applications*, Spring 1995, Volume 16, No. 1, pages 1–8.

[4] Wardrop, R.: "Student Sports Projects in a Statistics Course," An invited paper presented at the Joint Statistical Meetings, Section on Statistics in Sports, 1996.