Getting a Taste for Sampling: Using a Popsicle Stick Population to Illustrate Sampling Distributions

Robert J. Padgett and Kathryn A. Morris
Butler University
Indianapolis, IN 46208

Many texts fail to illustrate the critical role of sampling distributions in statistical inference. In our exercise to explore sampling distributions, students in basic and advanced courses use a population of 5000 popsicle sticks to learn about key statistical concepts related to sampling, sampling distributions, statistical power, and errors.

Purpose:

In our session we present a technique to introduce to students, in a tangible way, arguably the most central concept of inferential statistics -- the sampling distribution. Despite the centrality of the concept, few introductory statistics texts devote much attention to how sampling distributions provide the conceptual rational that underlies virtually every reject/fail to reject statistical decision. Texts that do present the concept usually do so by providing the mathematical derivations and descriptions of the properties of sampling distributions with little explanation of the relationship between the theoretical distribution and the process of sampling used in empirical research. We have developed a population of 5000 popsicle sticks that we use in basic and advanced statistics courses to demonstrate this and many other key statistical concepts.

In our courses, we typically utilize this population when we first introduce the concept of a sampling distribution, which enhances student learning by allowing them to actively select samples from the population and observe how the calculated sample statistics result in the values expected on the basis of mathematical theorems. Once students have had this direct experience with sampling, we believe it is easier for them to conceptually grasp sampling distributions created via computer simulation. For example, we initially build a frequency distribution of sample means by plotting the sample means from 20 samples selected from the population. Then, we move to computer simulations that create a sampling distribution by simulating repeating that same sampling procedure 20,000 times. Thus, we believe that a combination of hands-on sampling and computer simulation affords students with a strong foundation on which to build their knowledge of sampling distributions.

Example Applications of the Popsicle Stick Population
There are a variety of ways to use the popsicle stick population in class. Below are some ways that we have found to use the population to illustrate various statistical concepts.

1. **Create a sample and compute sample statistics:** Have students each draw a sample of 20 sticks, record the numbers and calculate the mean, standard deviation and other descriptive statistics. They can also use the numbers to create histograms and other graphic displays of data.
2. **Compute z-scores of individual data points in the population:** Use the population to calculate z-scores and the probabilities associated with sampling individual observation greater than, less than, or between some range of scores.
3. **Compute z-scores associated with particular sample means:** Have students calculate a z-test using their sample and construct a confidence interval for the true population mean.
4. **Demonstrate Type I error rates:** Have all students report their sample mean to the instructor and look for the number of Type I errors that occur. We reliably find that the Type I error rate of this procedure is right at stated alpha.

5. **Demonstrate how sample size affects standard error of the mean:** Have students draw 5 samples where the sample size is 5, and then 5 samples where the sample size is 15 and compute the mean of each sample as well as how much the means of each of the samples of 5 and 15 vary from each other. Have students compare that number to the theoretically expected standard error of the mean.
6. **Introduce t and F tests:** Use the population to illustrate drawing two samples from the population to introduce independent t-test, and paired t-test procedures. This same procedure can be used to illustrate within and between group variance with more than two groups (e.g., one-way ANOVA models).
7. **Illustrate correlation:** Have students draw paired samples and calculate the correlation between items. This process can be used to illustrate how the Person Product Moment Correlation captures the relationship between two variables and can be used to illustrate hypothesis testing with correlation (e.g., Ho: $\rho = 0$).
8. **Illustrate statistical power:** Have students assume that they are testing the null hypothesis Ho: μ=20 and then have them draw various samples from the population to see how many times they reject Ho based on the observed sample mean. We reliably find that the rate of rejecting Ho in this procedure is equal to the true power (1-B) of the test based on the sample size and alpha level.

Creating the Popsicle Stick Population:

Although it would be relatively easy to create a population of numbers printed on slips of paper, it was our goal to create a population of numbers using a durable product that is easy to handle and difficult to lose. We originally sought to create the population by writing numbers on poker chips. However, this proved to be too expensive. After investigating various options, we came across popsicle sticks, which are sold in bulk at arts and crafts stores for very minimal price. In fact, the total cost for 5000 popsicle sticks was approximately $35.

After purchasing the popsicle sticks, we created a program in SPSS to generate a normally distributed set of 5000 numbers having a particular mean and a particular standard deviation. Because we did not want students to consider the population to have any particular meaning, we specifically avoided using population parameters that are otherwise meaningful (e.g., intelligence test scores with a mean of 100 and a standard deviation of 15). Although we developed a population with a mean of 28 and a standard deviation of 5, any population parameters could be used.

Finally, we had student assistants write the data points on each popsicle stick, and we store the popsicle sticks in a plastic container large enough to allow manual mixing to ensure random sampling. Although we have not done so at this time, one could conceivably use the same popsicle sticks to create multiple populations by writing numbers in different locations along the sticks and/or using different colors of ink. Doing so would allow instructors to demonstrate sampling from normal versus non-normal distributions (e.g., skewed distributions or rectangular).

For further information about how to construct or use the “Popsicle Population” to illustrate statistical concepts please feel free to contact us via email. Robert J. Padgett may be contacted at rpadgett@butler.edu and Kathryn A. Morris may be contacted at kmorris@butler.edu