ILLUSTRATING INFERENCE IN AN INTRODUCTORY STATISTICS COURSE USING SIMULATION

Jeff Kollath
Department of Statistics
Oregon State University
Corvallis, OR, USA 97331
kollath@stat.oregonstate.edu

WHY SIMULATION?

A sampling distribution of a statistic is used in inference, but students often have a hard time grasping what a sampling distribution is, how it is used, and most importantly why it is necessary. Once inference is introduced, students have difficulty making the connection in how the sampling distribution is used. As an example, students are familiar with putting numbers into a formula to estimate the standard deviation of a sampling distribution, but often do not know what that number means or why it's necessary. Simulation can help with understanding how a sampling distribution is created.

Using simulations, students can gain an understanding of how the sampling distribution is used in the construction of a confidence interval for a parameter. But what is even more intuitive is how the sampling distribution is used in the calculation of a p-value. For an introductory student, the definition of a p-value is complex. It is often hard to connect the area under a curve with the definition of a p-value. However, when a sampling distribution is created through simulation under the null hypothesis, students can count the number of sample statistics "as or more extreme" than the one observed to calculate the p-value, which is much easier to understand.

HOW CAN SIMULATION BE INCORPORATED INTO AN INTRODUCTORY STATISTICS COURSE?

This poster presentation will show how simulation is used in my introductory courses. The two introductory courses I teach are upper division courses for students taking their first statistics course. The prerequisite is high school algebra with some basic statistics. The introductory sequence consists of two 10-week courses. In particular, this poster presentation will show or demonstrate the following:

- Present the course outline for my courses, which includes limited discussion of probability and introduces inference by creating a bootstrap distribution of sample means.
- Demonstrate different macros written in Minitab that are used in both introductory courses to find confidence intervals and p-values using bootstrap distributions
- Show that using a bootstrap distribution for confidence intervals is often the same or better at capturing the population parameter the percent of time corresponding to the confidence level than traditional "normal-based" confidence intervals.
- Illustrate that an "adjusted" bootstrap distribution created under the null hypothesis is often the same or better than traditional "normal-based" methods in terms of the desired probability of making a Type I Error.

In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in statistics education. Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9, July, 2014), Flagstaff, Arizona, USA.* Voorburg, The Netherlands: International Statistical Institute. iase-web.org