Sampling Distributions

Author: Rachel Hartnett, Oklahoma State University

Focus: The focal problem being is addressed is twofold: a) how are sampling distributions made and b) how does the spread of this distribution relate to the precision of a sample.

Overview: My swirl lesson is designed to help students with some of the most difficult, yet critical, concepts in unit one of my introductory biostats course: sampling distributions, and particularly null distributions. Students have difficulty in understanding the properties of sampling distributions (and don't know why we are even learning about them). So I've designed a supplemental lesson in order to reinforce the lecture. This swirl lesson will provide scaffolding for tricky programming functions, so students can focus on the parameters/conditions that are used to generate sampling (and later more specifically null) distributions. I expect students to know how to make a vector in R, how to interpret a histogram, and basic understanding of addition rules of probability.

Learning objectives: Learning objectives: 1) visualize how sampling distributions are made, 2) determine the key properties of the sampling distribution of a mean

Lesson sequence: Provide a numbered, ordered list of the activities within your swirl lesson. This list can be taken from step 4 in your initial lesson design, with any modifications that were introduced.

1. _A lecture the class period before the swirl lesson on sampling distributions
2. _the swirl lesson that visualizes probability, frequency, and sampling distributions of a mean
3. _Formative assessment within the swirl lesson of the properties of a sampling distribution of a mean and how that relates to a single sample
4. _Worksheet assessment outside of the swirl lesson that reinforces the questions asked during the swirl lesson

Pre-lesson activities: A lecture over sampling distributions that includes all of the content given via the swirl lesson. In addition, students took the swirl lessons "Basic Building Blocks" within the R programming course (found at https://github.com/swirldev/swirl_courses).

Post-lesson activities: Assessment of the swirl lesson includes an exercise that covers similar content, and a section in the first exam. The questions aimed at assessing knowledge of sampling distributions include the following:

For the next set of questions, using the following data:

[Data taken from Whitlock and Schulter, 2nd ed., example 3.2.] Running speeds (cm/s) of the male *Tidarren* spider: 1.25, 2.94, 2.38, 3.09, 3.41, 3.00, 2.31, 2.93, 2.98, 3.55

1. If you subsample four data points from this set of ten points, will you get the same average? Why or why not?

2. Take one subsample (n=4), record it below, and calculate the mean and standard deviation. What does the mean and standard deviation say about the data?

3. Define the "sampling distribution." How would you calculate the sampling distribution from the above dataset?

4. What does the spread of the sampling distribution tell you? What would happen to the spread of the sampling distribution if you increased your sample size?

Implementation notes: This lesson has not yet been implemented in a classroom. Any suggestions are welcome to rachel.hartnett@okstate.edu!

Because this is drawing a random sample and the sampling distribution of means is uniquely generated, there will be variation in the answers generated. E.g., the standard deviation of the sampling distribution of means won't exactly match the standard error of a single mean, etc. The formative assessment within the lesson asks for how close a value is relative to each other, but I would reinforce that all of these values are estimates, except when calculating directly from the population data.

Resources used: This is a lesson developed from the R code provided by Whitlock and Schulter. Code used to generate figures and samples are fully credited to their work on http://whitlockschluter.zoology.ubc.ca/wp-content/rcode/chap04.r