Null Distributions

Author: Rachel Hartnett, Oklahoma State University

Focus: The focal problem being is addressed is twofold: a) how are null distributions a special case of sampling distributions and b) given an observed dataset, how the test statistic, p-level, and hypothesis testing logic should be used to draw conclusions in inferential tests.

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: null distributions. Students have difficulty in understanding the properties of null 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 null distributions. In addition, it should provide students with step-by-step calculations of the test statistic, p-level, and then fail to reject/reject the null in order to ensure they are seeing each step occur rather than get the output all at once and not know how to interpret it (beyond IDing the P-level). 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. I also expect them to have the most difficulty with knowing that a null distribution is specific to your sample size, and how the test statistic relates to the p-level and alpha.

Learning objectives: Learning objectives: 1) see how null distributions are a specific type of sampling distributions and ID parameters to generate a null distribution for a test, 2) connect where the test statistic falls under the null to the p-level outputted by a test, as well as connecting that you can only make a statement about the null hypothesis (reject, fail to reject) because you are using null expectations in generating a sampling distribution.

Lesson sequence:

1. _A lecture the class period before the swirl lesson on hypothesis testing and null distributions
2. _The swirl lesson
3. _Formative assessment within the swirl lesson of the properties of a null distribution and how that relates to the test statistic and p-level in a binomial test

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), and “Sampling Distributions” found within the same course, “Distributions” , as this one.

Post-lesson activities: A Canvas assessment of the swirl lesson that includes an exercise that covers similar content, and a section in the first exam. Example questions include:

1. A _____ distribution reflects the distribution of a variable in a sample, while the _____ distribution gives the real distribution of the variable in the whole population. Often the _____ distribution is used to approximate the shape of this population distribution.

2. Circle one option to fill in the following sentences. The null hypothesis is framed to make a claim about a specific (sample statistic / population parameter). A null distribution is the (frequency/ probability /sampling) distribution of the test statistic, assuming the null is true. The null distribution assumes that (random /deterministic) processes are occurring. The (P-value/ significance level) is the probability of your observed data being unusual compared to a(n) (P-value/ significance level).

3. You poll a county by randomly sampling 25 citizens you find at grocery stores to see whether they would vote for Candidate A or Candidate B. You find that 15 out of 25 citizens say that they would vote for Candidate A. How would you generate the null distribution in this scenario?

4. How do we use the calculated test statistic in hypothesis testing to know when to reject the null hypothesis?

Implementation notes: This lesson was designed to take anywhere from 25-45 minutes, depending on experience using R. This can be implemented with the "Sampling Distributions" lesson, or with an accompanying exercise that assess concepts from the swirl lesson for a 75 minute class. This lesson has not yet been implemented in the classroom. Any suggestions are welcome to rachel.hartnett@okstate.edu!

Resources: 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/chap06.r