

Basic Statistics

Authors: Emily G. Weigel (Georgia Institute of Technology) and Alycia R. Lackey (Binghamton University)

Focus: The students will practice identifying the appropriate basic statistical tests when given a scenario and learn how to run and interpret those statistical tests in R.

Overview: This lesson centers around introducing 4 fundamental statistical tests used in biology: correlations, t-tests, ANOVAs, and X^2 (chi-square) tests. Students will first work through a worksheet to identify when each test should be conducted and what the appropriate research hypothesis would be for each experiment. Students then turn to swirl to dig into statistical hypotheses in the context of these scenarios to run and interpret statistical analyses in R.

Learning objectives:

1. Identify experimental variables and the appropriate basic statistical test for a given scenario
2. Practice coding elements necessary to run a statistical test in R
3. Interpret the output of a statistical test relative to a research question/hypothesis

Lesson sequence:

1. Introduce the basic statistical tests using the scenario packet, letting students work in groups (Note: an introduction to basic statistical parameters, like mean and distribution, may be needed as a prior step if these concepts are new to students)
2. Swirl lesson
 - a. Introduction
 - b. Basic Building Blocks
 - c. Looking at Data
 - d. Hypothesis Testing
3. Have the students use these skills to work on a new dataset of their choice (e.g., derived from in-class experiments, Dryad, etc.)

Pre-lesson activities: Students should be introduced to the basic statistical tests, as well as any necessarily basic descriptive statistics necessary to help them choose a test (e.g., defining a mean, distribution, etc.). Instructors should first focus on students understanding of why a test is applied over the specific values achieved, and should strongly consider using graphs alongside strict data values.

Post-lesson activities: For assessment, students should submit commented code from working through the swirl lesson, which can serve as a reference for future work. If students move on to analyzing a dataset of their choice using one of these tests, it is advised that instructors check in with students to briefly view code from the swirl lesson and the student interpretations of the statistical output before the students apply their code to a new dataset.

Implementation notes: This lesson is quite flexible and possible to do in stages across several class periods or as a series of homeworks. It is advised, however, to have discussions/check-ins with students strategically after they have completed the first 4 scenarios (to be sure the correct tests were chosen), the second 4 scenarios (to ensure the correct tests and variables could reliably be chosen again) and after completing swirl (to ensure the students have good notes for how to code and interpret statistical tests in R). It is advised that instructors not proceed to swirl/R until students first have a grasp on which tests to use under what scenarios. The scenarios take approximately 20 minutes to complete for introductory-level students; the swirl lesson is much more variable, but takes generally 40 minutes for a student taking good notes as they go. Encourage students to go slowly and keep a document or script open in which to code/take notes as they work; it will make generating their own R code easier later.