Practicing Statistics: Guided Investigations for the Second Course
Chapter 1: Randomization Tests and Nonparametric Tests
Instructor Notes

This chapter introduces the concept of statistical inference through a simulation study. Students are asked to conduct a randomization test to compare two treatments for a disease occurring in humans. Schistosomiasis (shis-tuh-soh-mahy-uh-sis) affects about 200 million people worldwide and can cause death, but more commonly results in chronic and debilitating symptoms, caused primarily by the body's immune reaction to parasite eggs lodged in the liver, spleen, and intestines.

R and Rstudio: If students are new to R and Rstudio, the following handout outlines some nice YouTube videos and short assignments to get students started with R: Introduction to Rstudio. Updated Chapter 1 R Instructions are also available.

PowerPoint Slides, Updated Datasets, Step-by-step software instructions and Instructors guides are available upon request. All the materials on the textbook CD are also at: http://www.pearsonhighered.com/mathstatsresources/ (just select K).

Below is a Chapter 1 syllabus appropriate for an introductory class, a full semester syllabus appropriate for a senior seminar or graduate level class is here: Advanced Syllabus

Day 1: Why Use Randomization Tests
Randomization tests, permutation tests, and bootstrap methods are quickly gaining popularity as methods for conducting statistical inference. Why? These nonparametric methods require fewer assumptions and provide results that are often more accurate than those from traditional techniques using well-known distributions (such as the normal, t-, or F-distribution). These methods are typically based on computer simulations instead of assumptions about distributions and thus are particularly useful when the sample data are skewed or when the sample size is small. In addition, nonparametric methods can be extended to other parameters of interest, such as the median, whereas the well-known parametric methods described in introductory statistics courses are often restricted to inference for the population mean.

1. **Watch the *VIDEO* *C1a: Why Randomization Tests*** (The same video with embedded quiz questions is available at http://www.screencast.com/users/Stat2labs/folders/Randomization%20Tests). This introductory video demonstrates how two-sample t-tests and nonparametric tests (such as randomization tests) can be used to compare two groups. It discusses why randomization tests are often preferred over traditional parametric tests.
2. **Read: Chapter 1.1** of the textbook (and work through the corresponding textbook questions: Questions 1 and 2)

Day 2: Conducting a Randomization Test by Hand

We begin this chapter by comparing two treatments for a potentially deadly disease called schistosomiasis (skis-tuh-soh-mahy'-uh-sis). We illustrate the basic concepts behind nonparametric methods by using randomization tests to provide an intuitive description of statistical inference and conduct a randomization test by hand

1. **Watch the VIDEO *C1b: Introduction to Randomization Tests*** (The same video with embedded quiz questions is available at http://www.screencast.com/users/Stat2labs/folders/Randomization%20Tests/media/737235ce-1b8b-46f3-af19-aae97ece746a)
   This video demonstrates what statistical software is doing when it calculates p-values through a randomization tests using the textbook example from a 2007 study on schistosomiasis.
2. **Read: Chapter 1.2** of the textbook (and work through the corresponding textbook questions: Questions 3-6)

Note: Clearly computer simulations are typically used to calculate the p-values in these tests. However, conceptual understanding appears to be much stronger when students spend 10-15 minutes actually attempting the randomization process by hand.

Day 3: Completing a Computer Simulation

1. **If using R, watch the VIDEO *C1 R: Using R to Conduct a Randomization Tests*** (A slightly longer video is at: http://www.screencast.com/users/Stat2labs/folders/R%20commands/media/18bc00d6-ddaa-4600-b51b-7862691c78ed). Note that the exact same instructions do work in both R and Rstudio. or
2. **If using Minitab, watch the VIDEO *C1e: Minitab Instructions for the Randomization Test*** (A longer, more course specific video is at: http://www.screencast.com/users/Stat2labs/folders/Randomization%20Tests/media/2f0e4d9d-1fd6-44a9-be1a-8e69944f6bdd).
3. Optional: You can also watch the slightly more advanced video on writing functions and scripts at: http://www.screencast.com/users/Stat2labs/folders/R%20commands/media/b835c47a-4b65-4f41-b273-46c0708afaa4
4. **Read: Chapter 1.3-1.5 (**and complete corresponding questions**).**
5. Submit solutions to **Questions C1.10, C1.12, C1.15, and C1.16.** The histogram in question 10 should be similar (but not identical) to Figure 1.2.

Note: **Question 10:**

The histogram is roughly normal, but because there are a limited number of possible outcomes (only 5 samples in each group), the histogram of (MeanCtl – MeanTrt) will never look perfectly normal. Students who run 10,000 iterations or less may find an asymmetric graph with one or two bars that looks high in their graph, however, theoretically we know this graph should be symmetric (if we conducted enough iterations). So in their explanation, don't focus on an anomaly that might have occurred in their graph, but discuss the long terms patterns in the graph. If they conducted the study again, we would not necessarily expect the same high bar in the same place. In the long term (100,000 or more iterations) we will see more and more symmetry, but we will not ever see normality.

Question 15: While the histogram for Question 15 tends to be roughly symmetric, it would be good to explain why we expect symmetry. Since every mouse has an equal chance of falling into either group, we can theoretically state that that if you conducted enough simulations it would look completely symmetric.

Question 16: Some students discuss how a one-sided test is a better choice than a 2-sided test. However, the key point to this question is that with the small sample sizes (only 5 observations each) the t-statistic may not follow the theoretical t-distribution. The previous questions verified this because the p-value of around .055 is much more accurate than the 0.022 value provided by the t-test.

Discussion Questions (completed in class on on-line): As we complete this chapter. Have students answer any two of the following discussion questions.

- What is the most challenging concept for you in understanding the randomization test for the schistosomiasis study?
- K11777 only reduced the number of worms, shouldn't we just stick with praziquantel?
- Why should boxplots or other graphical techniques be used to visualize data before a parametric test is conducted?
- What is the difference between a random sample and a randomized experiment?
- Would you trust the results of the study if the researchers had haphazardly pulled the mice from a cage and assigned the 1st five to the treatment and the last five to the control? Why or Why not?

Day 4: Multiple Comparisons
Read: **Chapter 1.13** and submit **questions C1: 29-35**.

Day 5/6: **Power and p-values** and corresponding video: ***The Limitations of p-values***
(https://www.youtube.com/watch?v=pr3CrXgvWeE)
At this time, the above game-based lab can be used to enhance students understanding of when p-values can be misleading.

Additional Class Discussion:

- Make sure to emphasize the importance of a simple random sample when conducting a hypothesis test, not just having a large enough sample size.
- It is often interesting to discuss why small samples are typically used in medical testing on animals. The data in this chapter represent one phase of the schistosomiasis study. Ten female laboratory mice and ten male laboratory mice were deliberately infected

with the schistosome parasite. At the end of the study, the mice were euthanized to count both the number of eggs and the number of worms in the mice livers.

- Many students have a tendency to incorrectly believe that the t-test is always appropriate when comparing the means of two populations (everyone gets exactly the same p-value on textbook data with parametric tests). This lab demonstrates how p-values from permutation/randomization tests are more accurate than the traditional two-sample t-test. In addition to reviewing hypothesis testing techniques, this chapter emphasizes the importance of checking assumptions.
- Several people come to class with the assumption that all nonparametric tests tend to have lower power than parametric tests. While this is true for tests such as the Kruskal-Wallis test, Sign Rank test, or Wilcoxon rank sum test, it is important to understand that the randomization tests and permutation tests do not tend to have lower power. In fact, as section 1.10 points out, Ronald Fisher describes the t-test as an approximation to the typically more accurate randomization test.