An Active Learning In-Class Demonstration of Good Experimental Design

C. M. Anderson-Cook and Sundar Dorai-Raj
Virginia Polytechnic Institute and State University

Journal of Statistics Education Volume 9, Number 1 (2001)

Copyright © 2001 by C. M. Anderson-Cook and Sundar Dorai-Raj, all rights reserved. This text may be freely shared among individuals, but it may not be republished in any medium without express written consent from the authors and advance notification of the editor.

Key Words: Blocking; Internet; Java applets; Paired data; Randomization; Teaching; Two-sample t-tests.

Abstract

This article presents an active learning demonstration available on the Internet using Java applets to show a poorly designed experiment and then subsequently a well-designed experiment. The activity involves student participation and data collection. It demonstrates the concepts of randomization and blocking, as well as the need to carefully consider the objective of a study and how well the data collected answer the question of interest. The proposed exercise takes approximately 50 minutes of lecture time and helps to solidify these essential statistical concepts in a visual and memorable way. A variation of the activity could extend the presentation to 75 minutes. Students have reacted positively to the exercise.

1. Introduction

This one-class exercise is proposed to give students in introductory statistics courses or a first class in experimental design some hands-on experience with designing a real experiment, the data collection process, and simple analysis of the data. The importance of students experiencing data collection has been well documented by Hunter (1976), Hogg (1991) and Mackisack (1994). However, in introductory courses with a large number of topics to discuss, it is not always possible to include the data collection projects suggested by Fillebrown (1994), Anderson-Cook (1998), or Short and Pigeon (1998). Online Web-based materials, including West and Ogden (1998) and Vestac (http://www.kuleuven.ac.be/ucs/java/) tend to focus more on statistical concepts or the analysis aspect of experimental data, rather than its collection. Links to these and other Java-based education programs can be found under the "Information and Help" links at http://www.stat.vt.edu/~sundar/java/applets/.

Our experiment involves comparing reaction and execution times for dominant and non-dominant hands using a computer mouse. The activity consists of running two experiments to collect data. It begins with a poorly planned experiment, a discussion of problems with the first experiment, and then a follow-up experiment where statistical and design issues are more appropriately handled. Design issues such as randomization and reduction in experimental error arise throughout the entire exercise. To run the experiment in class requires access to a computer with Internet capability, a Java-compatible Web browser, and the ability for the class to see the computer screen through the use of a projection unit or by gathering around the computer. An executable file containing all pertinent files and directories can be downloaded from http://www.stat.vt.edu/~sundar/java/ExpDesign/ for those who do not have classroom Internet access.

The original motivation for designing this activity was to give students some first-hand experience with data collection. Experimental design texts such as Box, Hunter and Hunter (1978), Montgomery (1991) and even some introductory statistics texts such as Wardrop (1995) discuss the importance of including issues related to decision-making and physical details of data collection in statistics courses. Exposure to experimental design and the issues of randomization and blocking become more interesting when properly motivated by a concrete problem.

It is not uncommon for students to understand the theory behind blocking and randomization as important statistical ideas while taking an introductory statistics course, and yet not to incorporate them into their own data collection practices when they become researchers in their own discipline. Angelo (1993) emphasizes that student learning increases dramatically if they are actively engaged in a demonstration of important concepts. As one former student, who then became a consulting client when designing his senior biology project said, "Oh, I didn’t think I needed to randomize – that seemed like something only statisticians would do. That ‘sticking numbers in a hat thing’ just did not seem scientific.” Collecting the data in this in-class activity has made the notions of
blocking and randomization seem more applicable to real world situations.

The demonstration consists of four phases. First, as the students walk into the class, data is collected from the poorly designed experiment using a laptop and projection unit. This animated activity at the start of the class is an excellent enthusiasm generator. Second, we discuss the goal of the experiment and the procedure used for data collection. Students are then asked to comment on some of the problems with the design, and how they could be eliminated with a more carefully thought-out and better-implemented experiment. Third, they rerun the experiment with the suggested improvements. The final phase, which can either be done in class or as an exercise for the students between classes, involves analyzing the data obtained from both experiments and comparing the results.

2. Details of the Exercise

This exercise should be run after students have been exposed to the concepts of randomization, replication and blocking, as well as the appropriate analyses associated with two-sample independent samples and two-sample paired samples.

2.1 Collection of Data from the Poorly Designed Experiment: Phase 1

Figure 1 shows a snapshot of the program used to collect data from the poorly designed experiment to compare differences in mean reaction and execution times between hands. The design suffers from increased experimental error by the placement of the start button as well as the layout of the squares. The Java applet can be found at http://ww2.amstat.org/publications/jse/java/v9n1/anderson-cook/BadExpDesignApplet.html. This program should be loaded as the students enter the class, and 16 to 20 students should be asked to come up to the computer. Each student is asked to click the "Start" button, wait for the 3-second delay, and then move to click the mouse on the square that changes color with the hand specified. Each student should be able to complete the trial 15 to 25 seconds. With the class able to watch the data collection, there is plenty of interest in the activity and students are usually eager to participate.

Examining Figure 1 more closely, we see that there are nine squares, each of which could be randomly highlighted for a given trial of the experiment. Students click on the "Start" button, and between the time in seconds and the "Start" button, the student is instructed which hand to use for the experiment. The red and green circles above the time illuminate sequentially to count down the three seconds until the square changes color. Once the student moves the mouse to the correct square and clicks anywhere inside the box, the timer stops and the elapsed time is displayed in the appropriate column below. After the first student, the "Reset" button is pressed and the procedure is ready for a new student. Once all of the students in the experiment have participated, the "Compile Info" button is pressed to give a summary of the data collected in the "Results" box in the top right corner. To save the data into a text file for later analysis, the "Save Data" button is used.

![Figure 1. The Poorly Designed Experiment.](image-url)
2.2 Discussion and Improvement of Experimental Design: Phase 2

The poorly designed experiment is supposed to be a reasonable first proposal for a data collection scheme by someone without much knowledge of statistically important concepts. The goal is to determine if there are differences in mean reaction and execution times between hands manipulating a computer mouse. In leading the discussion, the objective is to help the students identify shortcomings of the current experiment, and then to suggest improvements which could be implemented in the subsequent experiment. The improvements will fit into several categories, including more precisely defining the objective of the study, reconfiguring the experiment to reduce experimental error, and incorporating statistical concepts like randomization and blocking. Students will typically suggest many of the required improvements to the experiment either on their own or with some prompting toward a particular category of improvement to consider. Students may also suggest some improvements that have not been incorporated into the new experiment. This does not pose a problem, as the second experiment can be packaged instead as just an improved version of the experiment, not as an optimal one.

One of the key ideas that should be extracted from the discussion is that the notion of right and left handedness does not match exactly with dominant and non-dominant handedness for a segment of the population (namely the left-handed students). It does not make sense to group these results without careful consideration. Possible solutions may include restricting the experiment to right-handed people only, having the left-handed people use their left-hand for their dominant hand, or having the hand usually used for computer mouse manipulation declared the dominant hand. Any of these suggestions will work for the subsequent experiment, and the discussion can lead to the understanding that each potential solution will answer a slightly different question. The question of careful formulation of the goal of an experiment is an important one, yet one that could be overlooked in discussions of experimental design in statistics classes. An additional issue to be considered is the placement of the mouse, which if stationed in the traditional location to the right of the keyboard may compromise results. Several solutions to this may be possible, including centering the mouse below the keyboard or using the built-in mouse on many laptops that is located in the middle of the keyboard.

Reconfiguring the experiment to establish a reduction of experimental error is another area for discussion. One alternative to consider is to allow a practice attempt for the participants, so that the experiment is not measuring the learning curve of the students implementing the instructions, but rather their actual reaction and execution times. The first time through the experiment, there is usually at least one student who does not understand the procedure completely, and as a result ends up with a very large reaction time (an outlier). By allowing the students to take a practice attempt, the experiment may more accurately measure the quantity of interest. There could be some discussion of a protocol given to the participants, to ensure that all of them understand and follow essentially the same procedure. It is also helpful to redesign the appearance of the experiment to make all of the boxes equidistant from the position of the mouse at the start of timing. As the experiment is currently arranged, we would expect that times to move to boxes 3, 6 and 9 would be shorter than to boxes 1, 4 and 7, simply because of the distances moved. Moving the starting position to the center of the squares should make the times more consistent and hence decrease variability of the response times and reduce experimental error.

The statistical issues to be incorporated into the experiment are randomization and blocking. A brief discussion is helpful about why choosing the first students through the door may not be a good way of selecting participants in the study. The selection of the students could be randomized, to include both the "slow-to-class" and the "quick-to-class" students. Blocking is an effective method for extracting more precise information from the experiment. Since the goal of the experiment is to find the difference in the mean response times between dominant and non-dominant hands, obtaining two observations from each participant, one for each hand, is an effective way of collecting data. Reviewing the advantages of paired data over two separate samples is easily demonstrated, since students are intuitively aware that observations from the same student are more likely to be similar than observations from different students. Once the idea of two observations per participant has been introduced, it is important to consider randomizing the order of the runs for each person. Frequently, the discussion of various improvements to the experiment involves discussion among a large number of students, and generates considerable enthusiasm and creative thinking.

These key features have been changed from the experiment in Figure 1 to the new experiment in Figure 2:

- Terminology has been changed from left and right hands to non-dominant and dominant hands
- The "Start" button is now equally distanced from all eight of the boxes removing the additional source of variation associated with the different distances to travel with the mouse.
- Students now run the experiment twice each with the order of the hands has been randomized for each pair of observations.
- The students for the experiment have been randomly sampled from the entire class.
After the discussion of issues and ideas is completed, the new experiment shown in Figure 2 can be shown to the class and can be run in the next phase. Table 1 lists the differences between the two experiments and the problems associated with running the poorly designed experiment. The Java applet for the new experiment can be found at http://ww2.amstat.org/publications/jse/java/v9n1/anderson-cook/GoodExpDesignApplet.html.

### Table 1. Differences Between the Two Experiments.

|                  | Experiment 1 (Poor)                  | Experiment 2 (Improved)                  | Problem |
|------------------|--------------------------------------|------------------------------------------|---------|
| **Layout**       | 3 x 3 grid.                          | Equidistant from center.                | a       |
| **Start Button** | Right of grid.                        | Center of squares.                      | a       |
| **Hand Labels**  | Right/Left.                           | Dominant/Non-Dominant.                  | a       |
| **Randomization**| None                                  | Randomizes which hand goes first.       | b       |
| **Analysis**     | Two independent sample t-test (Pooled variance). | Paired t-test.                          | N/A     |
| **Run-order**    | Haphazard sample.                     | Simple random sample.                   | b       |
| **Figure**       | Figure 1.                             | Figure 2.                                | N/A     |

Problem Codes: a = Increases Experimental Error, b = Fundamental Design Flaw.

### 2.3 Collection of Data with the Improved Experiment: Phase 3

Collecting the new data with the revised experiment involves randomly selecting students from the class (or from among the right-handed people if this is the specific question of interest). This can be done by numbering the students from 1 to $n$, and then using the function displayed in Figure 3 and available on the Web page http://ww2.amstat.org/publications/jse/java/v9n1/anderson-cook/selection/ to generate a list of randomly selected numbers. Once the participants are selected, the protocol developed in phase 2
would be outlined (with the students possibly each taking a practice turn with the mouse to get the feeling for running the experiment). The precise statement of the protocol is an important aspect of formalizing the experiment, and helps to reinforce the idea of consistency across participants. To keep the data collection phase of the experiment to a minimal amount of time, it is recommended that 8 to 10 students participate in the experiment. Other students are encouraged to verify that the protocol is being closely followed.

![Randomization Software for Collecting Data](image)

### Step 1.

**How many students are in the class?**

[10]

**How many participants?**

[10]

Start Selection

Start Applet

### Step 2.

**Select an option.**

Randomize

Restart

### Step 3.

**Run the applet using the following randomized order of the participants.**

| Order | Student # |
|-------|-----------|
| 1     | 10        |
| 2     | 2         |
| 3     | 5         |
| 4     | 3         |
| 5     | 1         |
| 6     | 6         |
| 7     | 7         |
| 8     | 4         |
| 9     | 9         |
| 10    | 8         |

**Students = 10**

**Participants = 10**

### Step 4.

**Press the button to start the Good Experimental Design Applet.**

Note: Pressing the button below will load the applet in the current window and open a new window containing the randomized table from Step 3.

Start Applet

---

### 2.4 Analysis of the Data from Experiments: Phase 4

While the data is being collected for the second experiment, discussion can begin about how the data from the two experiments should be analyzed. For each of the experiments, the "Save Data" button allows the data to be saved to a comma-delimited file. If students are on a listserv or an e-mail distribution list, the data can be mailed to them. Otherwise, the students can copy down the important numbers and perform the analysis by hand or enter it into a statistical computer package. The "Results" box for each experiment (in the top right corner) gives the test statistics with the appropriate degrees of freedom. Figure 4 shows the window that allows for the copying of the data to a file.
In the first experiment, the conclusions may be that there are problems associated with the design, but that it is still possible to see if there are any differences between the mean responses of the two groups. The data have intentionally been collected to look somewhat paired, although there is no reason to suspect that pairing the initial data is sensible or advantageous. The applet computes the sample statistic for conducting a pooled two-sample t-test, although there are other tests that could be performed. If the latter test is conducted an initial informal check should be made to verify that the data are normally distributed.

The second experiment with two observations per participant should be analyzed with the paired t-test. Again, the results from pressing the "Compile" button can be verified by hand or computer calculation. We omit p-values from the summary, as this allows students to draw their own conclusions about the results of the two experiments. Depending on the timing of the class and the comfort level of the students with the calculations, the conclusions can be discussed at the end of a single class period or given as an exercise to be completed before the next class. In either case a follow-up discussion should be held to interpret the results of the analyses.

3. Results and Comments

This experiment has now been run several times with introductory statistics classes at Virginia Tech. In every case, the results of the analysis have failed to show any significant difference in mean response between hands for the first experiment, but a significant difference in the mean response between dominant and non-dominant hands for the paired data. This simple display of the reduced experimental error is an important lesson for good experimentation practice.

3.1 Variations on Running the Experiment

If additional time is available, an additional phase can be added before the second run of the experiment. This would involve running the first experiment again, now with each participant doing each of the two hands. In this case, as with the first experimental set-up, the right hand is always run first. This is a particularly useful experiment to run with a more advanced statistical experimental design class, since the lack of randomization of run order causes the confounding of the dominant and non-dominant hands with the learning curve of the participants, and the resulting test will usually fail to find a significant result. If this phase is run, the three experiments can be compared to illustrate the problems that confounding can cause in being able to detect an existing difference. When this was done with a class at Virginia Tech, the paired t-test with non-randomized data (and no practice) did not show a significant difference in the mean response between the dominant and non-dominant hands.

3.2 Student Reaction

Student reaction to the in-class activity has been generally positive. On the day of the demonstration there is typically a noticeable increase in the level of attentiveness, enthusiasm and participation. On final evaluations for the course, this activity is mentioned as one of the highlights of the course. On questions about experimental design and the importance of randomization and blocking.
student understanding appears to have improved and they are better able to articulate these concepts on exams given after the demonstration.

3.3 Advantages and Disadvantages of Using Java

We chose Java to create this exercise for two main reasons. First, the object-oriented environment allowed us to create subclasses that are shared by both the poorly designed and well-designed experiment applets. Java yields reusable code and smaller file sizes, which subsequently produce a faster loading time. This is crucial for those who run the applets directly from the Internet. The second reason for choosing Java is its portability through the Internet. The applets presented here have been thoroughly tested on the Windows and UNIX platforms using both Netscape 4.08 and Internet Explorer 5.0. The Macintosh version of Netscape will not support the Java applets until Netscape updates their browser to the latest Java virtual machine. The Macintosh version of Explorer 5.0 seems to have no problems. More information on the Macintosh can be found at [http://www.apple.com/](http://www.apple.com/). The only additional software required to use these exercises is a Java-interpreter, which comes standard with most Web browsers.

4. Conclusions

This simple interactive exercise can be easily integrated into any introductory statistics or first experimental design course. Access to a computer with Internet capability and the ability to allow students to view the screen are the only requirements for its implementation. The exercise is visual and active, and students can visualize concepts such as randomization and blocking while gaining experience with implementing these ideas. This activity could also be a launching point for a broad discussion about a variety of other statistical topics. Determining adequate sample size and power are two topics that could be introduced and taught as a follow-up. More complex experiments involving other factors such as gender, computer experience or type of mouse could also be considered.

References

Anderson-Cook, C. M. (1998), "Designing a First Experiment: A Project for Design of Experiment Courses," *The American Statistician*, 52, 338-342.

Angelo, T.A. (1993), "A Teacher’s Dozen: Fourteen General Research-Based Principles for Improving Higher Learning in our Classrooms," *American Association for Higher Education Bulletin*, 45, 3-13.

Box, G.E.P., Hunter, W.G. and Hunter, J.S. (1978), *Statistics for Experimenters*, New York: John Wiley & Sons, Inc.

Fillebrown, S. (1994), "Using Projects in an Elementary Statistics Course for Non-Science Majors," *Journal of Statistics Education* [Online], 2(2). ([http://ww2.amstat.org/publications/jse/v2n2/fillebrown.html](http://ww2.amstat.org/publications/jse/v2n2/fillebrown.html))

Hogg, R.V. (1991), "Statistical Education: Improvements are Badly Needed," *The American Statistician*, 45, 342-343.

Hunter, W.G. (1976), "Some Ideas about Teaching Design of Experiments, with 2^5 Examples of Experiments Conducted by Students," *The American Statistician*, 31, 12-20.

Mackisack, M. (1994), "What is the Use of Experiments Conducted by Statistics Students?" *Journal of Statistics Education* [Online], 2(1). ([http://ww2.amstat.org/publications/jse/v2n1/mackisack.html](http://ww2.amstat.org/publications/jse/v2n1/mackisack.html))

Montgomery, D.C. (1991), *Design and Analysis of Experiments* (3rd ed.), New York: John Wiley & Sons, Inc.

Short, T.H. and Pigeon, J.G. (1998), "Protocols and Pilots Studies: Taking Data Collection Projects Seriously," *Journal of Statistics Education* [Online], 6(1). ([http://ww2.amstat.org/publications/jse/v6n1/short.html](http://ww2.amstat.org/publications/jse/v6n1/short.html))

Waldrop, R.L. (1995), *Statistics: Learning in the Presence of Variation*, Dubuque, IA: William C. Brown Publishers.

West, R.W. and Ogden, R.T. (1998), "Interactive Demonstrations for Statistics Education on the World Wide Web," *Journal of Statistics Education* [Online], 6(3). ([http://ww2.amstat.org/publications/jse/v6n3/west.html](http://ww2.amstat.org/publications/jse/v6n3/west.html))

C. M. Anderson-Cook
Department of Statistics
Virginia Polytechnic Institute and State University
Blacksburg, VA 24061-0439
