

Small Student Projects in an Introductory Statistics Course

Robert L. Wardrop*
Department of Statistics
University of Wisconsin-Madison

July 3, 1999

1 Introduction

The key to effective public speaking, I have been told, is to begin with a funny story. Thus, I will begin this article with a story.

There was a one-room school house in a remote rural area. Periodically a regional supervisor would visit the school to evaluate the performance of its teacher. On one visit the supervisor viewed a unit on the addition of fractions. The teacher wrote on the board, $2/3 + 4/5 = 6/8$. After school that day the supervisor spent a great deal of time ensuring that the teacher understood the correct way to add fractions.

On the next visit to the school the supervisor was dismayed to find the teacher again was using the incorrect method of adding fractions. Confronted later in the day, the teacher replied, "I tried your method, but mine is easier to teach."

The method's benefit to the teacher is, of course, overwhelmed by its disservice to the students. I will share briefly some ideas I have on what makes a statistics course good or bad. In particular, I will argue that on certain occasions a teacher should opt for a more difficult way to teach statistics.

My ideas can be summarized quite well by the paradigm: be willing, not willful. Below is an example of a willful statistics exercise.

Sally has a random sample of size $n = 10$ from a normal population with unknown mean and standard deviation. Use Sally's data to construct a 95 percent confidence interval for the mean of the population.

This exercise is willful because its author tells the student exactly what to assume in order to complete it. This exercise is easy for the instructor because there is a known unique correct answer.

A willing exercise would provide a description of the scientific problem that motivates Sally's work. It would carefully describe how Sally obtained her data. Finally, a willing exercise would ask the student to explain what can be learned from Sally's data. The willful exercise amounts to teaching the student to turn a crank or push a button. The willing exercise forces the student to think; to draw upon ideas presented in the class along with other knowledge the student might have and combine these in a creative way. The willing exercise is difficult for the instructor because he or she does not possess the unique correct answer. I learned

*Robert Wardrop is Professor, Department of Statistics, University of Wisconsin-Madison, 1210 West Dayton Street, Madison, WI 53706.

long ago that if I assign willing exercises, I must check my ego at the door—frequently one of my students finds a clever and creative answer that I missed.

Certainly, some willful exercises have a place in the best of courses. My complaint is with courses whose only goal is to prepare students to satisfactorily complete willful exercises. There is a great temptation to omit any willing exercises from a course. *Willing exercises take time*. As long as there are powerful people who judge a statistics course by “How many topics are covered,” teachers will face pressure to omit willing exercises, leaving more time to cover more topics.

I consider the subject of statistics to be a collection of concepts, principles, and methods that help scientists learn about the world. (I define a scientist as any person who wants to understand better his or her natural, political and social worlds.) Mathematics is a critical component of statistics, and, I believe, all other things being equal, the person who is the better mathematician will be the better statistician. If my view has any validity, should not our introductory statistics courses make some effort to help students become better scientists? For example, the first day of class I announce that the fundamental goal of my course is:

To enable students to discover that statistics can be an important tool in daily life.

I want my students to learn that statistics is useful because it can help *them* learn about things that *they* find interesting.

My fundamental goal has an impact on various aspects of my course: the choice of text, the material presented, the methods of presenting the material, and the methods of evaluating the student’s learning. In this paper I will limit attention to one aspect of my course: my use of small student projects. See [2], [3], [4], and [5] for related work. As will become immediately obvious in Section 3, student projects can give rise only to willing exercises.

2 Limitations

Beginning in the next section, I will present several examples of student projects from my classes. As will become obvious, my students deserve all the credit for the creativity and cleverness their work demonstrates. My main contribution is that I assign this work to them. I believe that an ideal introductory class would require each student to do several projects. Limits on my time, however, require that I assign only two projects in my course.

Based on my teaching experiences, I fervently believe that a student learns more by analyzing data that he or she has collected, than by analyzing data collected by another. This discovery did not surprise me. What did surprise me is that most of my students prefer studying other students’ projects to “professional” research that I present. I do not know why this is true; I conjecture that there is less variation in interests within students than there is between the students and me. This is the first of two reasons I use numerous student projects as examples in lecture—my students find them to be interesting. The second reason is that I learned in the early days of assigning projects that few students are creative without models of good projects, but almost every student is creative when provided models of good projects. (This partly explains why this paper contains numerous projects; if you choose to assign projects you might want to make the project topics below available to your students.)

Figure 1, taken from [1], displays four possible study designs and notes their allowable inference(s), if any. Units can be selected at random or not at random, and can be allocated to groups by randomization or not. If units are selected at random, then inferences to populations are valid. If units are allocated to groups by randomization, then causal inferences are valid. All introductory Statistics texts present inferences to populations, but relatively few discuss causal inferences. This is unfortunate. Arguably a key component

Allocation of Units to Groups

<u>Selection of Units</u>	<u>By Randomization</u>	<u>Not by Randomization</u>
At Random	A random sample is selected from one population; units are then randomly assigned to different treatment groups	Random samples are selected from existing distinct populations.
Not at Random	A group of study units is found; units are then randomly assigned to treatment groups.	Collections of available units from distinct groups are examined.

Figure 1: Statistical inferences permitted by study designs (from *The Statistical Sleuth* by Ramsey and Schafer). For the designs in the first row, inferences to populations can be drawn; for the designs in the first column, causal inferences can be drawn.

of statistical literacy is the ability to distinguish between causal and noncausal links between variables. It is difficult to distinguish a difference if one is never taught that such differences exist.

I want my students to be able to perform valid inference, but I do not want them to expend the time and effort necessary to obtain random samples. Hence, my students “work in” the lower left entry in Figure 1; they can draw causal inferences, but not population inferences. I refer to such inference as *randomization-based inference*. For details on randomization-based inference see [3], [4], and [5].

3 Eleven Projects with a Dichotomous Response

For many of my students, their daily life includes working to obtain enough money to continue school, and these students are very interested in ways to earn more money. The first four projects below were executed on-the-job. (Note: My students have given me permission to use their work and I will identify each project with the first name(s) of its author(s).)

1. **Lori** worked as a waitress and wondered whether suggesting a specific appetizer upon greeting her customers would lead to an increase in the sales of appetizers. Her data, though not statistically significant, ran counter to her belief—mentioning a specific appetizer decreased the sales of all appetizers!
2. **Nell** worked at a coffee cart which offered two sizes, small and large. Of course, some customers specified the size when they ordered, but Nell was interested in those who simply ordered coffee. She found that the question, “Would you like a large?” was statistically significantly superior to the question, “Would you like a small or a large?” at eliciting the purchase of a large cup of coffee.
3. **Andre**’s work problem was similar to Nell’s. He sold ice cream which could be served in a plain or a waffle cone. His store made a much larger profit on the waffle cones. For customers who did not specify the cone type in their order, Andre found that the question, “Would you like a plain cone or a

homemade waffle cone?” elicited significantly more sales of waffle cones than did the same question with the adjective “homemade” deleted.

4. **Mary’s** job duties included the purchase of used compact discs from customers. Mary wondered which of the following statements to the customer would be more effective.
 - I will give you \$X.
 - How does \$X sound?

(The value of X was appropriate for the number and quality of the compact discs offered for sale.) Mary found that the first statement performed better (at getting an acceptance of the first offer) than the second, but that the difference barely missed achieving statistical significance.

The following project was motivated by the experiences of its author and many of his friends.

5. **Tuan** was interested in the problems an international student faces at the University of Wisconsin–Madison. He showed 25 graduate students part of an essay he said was written by “Jack McConnell, a student from Iowa.” He showed 25 other college graduates the *same essay*, but said it was written by “Hsiao-Ping Zhang, an international student from China.” When asked, “Do you detect any grammatical errors in this passage?” 64 percent who had read the Chinese student’s essay said yes, compared to only 20 percent who had read the Iowa student’s essay! This huge difference is highly statistically significant.
6. **Ruth** visited a minimum-security federal prison camp to obtain her subjects, first-time nonviolent offenders. The first version of her question read,

The prison is beginning a program in which inmates have the opportunity to volunteer for community service with developmentally disabled adults. Inmates who volunteer will receive a sentence reduction. Would you participate?

The second version was the same, except that there was no mention of sentence reduction. Ruth was surprised when her data revealed that the second version received a much higher proportion of yes responses than the first version, but the difference did not quite achieve statistical significance.

In the above studies, the experimental units are people. More often, my students choose units that are trials. The remaining examples are of this latter type. Frequently, my students choose to use statistics to investigate longstanding hobbies or pastimes.

7. **Erin** began her project with the following statement.

Ice-skating has been a part of my life for 12 years and one thing that stands out in my mind . . . were the numerous arguments I would stubbornly have with my coach.

Erin’s trial was an attempt at an axel (a figure skating jump with one and one-half completed turns). Erin found that she was highly statistically significantly better at completing an axel off her right foot than an axel off her left foot.

8. **Kathy** found that she was much better at successfully completing a cartwheel if she led with her right hand rather than her left.
9. When signing the alphabet, it is standard to use one hand as the primary hand and the other as the assister. **Diana** found that choosing her right hand as primary made her a better signer than when her left hand was primary, but the difference just failed to achieve statistical significance.

10. **Lisa S.** found that she was a much better lacrosse shooter with her right hand than with her left.
11. **Mike** enjoyed riding his personal water craft very fast and making sharp turns. His pleasure was diminished, however, whenever he would be thrown from the craft and land 30 feet away from it. Mike found that he was much better at staying on-board if he turned left than if he turned right.

In the next section I will present projects with a numerical response. Before turning to them, I want to share a bit of what my students and I have learned from projects like those above.

Over the past few years I have graded thousands of student projects. A small number of the project-topics appears to have been chosen to complete the assignment with as little thought and effort as possible. For example, a study of the effect of the hand used on the outcome of the toss of a coin would be of this type. I make it clear to my students that it is not necessary for me to consider the topic to be interesting; it is necessary, however, that the project report convincingly explains why the topic is of interest to the student.

The great majority of the projects are creative and interesting, and not qualitatively different from the ones described above. Students show great interest in studying ambidexterity, and various sports and games, especially darts, golf, basketball, archery and sharpshooting.

Analyzing their own data makes students appreciate the importance of the P-value. Obtaining a very small P-value seems to make them think, “Yes! I have achieved results that standard scientific practice says are real!” When a difference that is large enough to be important is not statistically significant, most students realize in a very personal way the advantages of collecting more data. Unfortunately, some students misinterpret P-values larger than 0.05 as *proving* that the treatments are identical.

I learn more about what my students really understand by reading what they write freely, instead of relying strictly on responses to exam questions. Sometimes they display ignorance or misconceptions in areas I might never think of examining. Also, occasionally a student will make it quite clear that even though he or she understands what the data and P-value signify, no study is going to change the student’s view of the world! For example, a male student’s data revealed that male subjects lied about their age more than female subjects did, but the difference was not statistically significant (assuming random samples, which was invalid). The researcher wrote, “I don’t care what the data say, we all know that women lie about their ages more than men do.”

4 Thirteen Projects with a Numerical Response

One of the pleasures of a numerical response is that we can draw pictures of our data and see what the pictures reveal. Unfortunately, space limitations prevent me from drawing pictures for all of the projects of this section.

12. Sara enjoyed playing golf. Like many novice golfers, Sara wondered whether she was more effective at hitting a golf ball from a fairway lie with a 3 wood or with a 3 iron.
13. Brian was a student in the ROTC program. As part of his training, Brian was required to run in combat boots and in jungle boots. Brian wondered whether the type of boot affected his running.

Figure 2 presents stem-and-leaf plots of the distance, in yards, that Sara hit a golf ball, by club. Sara’s data with the 3 wood could be described as moderately skewed to the left, whereas her data with the 3 iron is not so well-behaved. I would argue that there are three distinct groups of observations plus a small outlier. The reader, of course, can reasonably disagree with my interpretation.

Figure 3 presents dot plots of the time, in seconds, Brian required to run one mile, by footwear. Although the distributions overlap, it is clear (descriptively) that Brian ran faster while wearing jungle boots than he

	Three-Wood		Three-Iron
2	2	2	7
3	28	3	
4		4	
5	68	5	23789
6		6	888
7	7	7	
8	1	8	248
9	39	9	22227789
10	111477899	10	01577789
11	0134568	11	068
12	22778889	12	7
13	11799	13	22667899
14	07		

Figure 2: Stem-and-leaf Plots of the Distance, in Yards, Sara Hit a Golf Ball, by Type of Club.

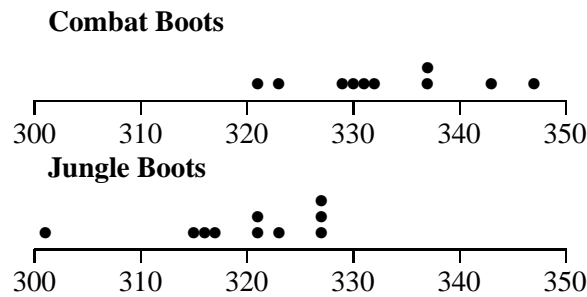


Figure 3: Dot Plots of the Time, in Seconds, Required for Brian to Run One Mile, by Footwear.

did while wearing combat boots. The distribution of the combat boot data is approximately symmetric, but for the jungle boots it is skewed to the left with a small outlier.

It is my experience that students find examining pictures of data to be more interesting, and they are better at it, when they have a personal interest in the data. In addition, typically it is easier and more meaningful to compare and contrast two related pictures than to evaluate one picture in isolation.

Brian's mean and standard deviation are 333.0 and 8.18 wearing combat boots, and 319.5 and 7.93 wearing jungle boots. Sara's mean and standard deviation are 106.87 and 29.87 with the three-wood, and 98.18 and 28.33 with the three-iron. Either Brian's or Sara's data can be used to illustrate various inference formulas. Notice that exercises of this type would naturally be of the willing variety. These data provide excellent opportunities for a consideration of issues like robustness. For these data the randomization distribution and t-distribution procedures give qualitatively the same answers. For Brian, the jungle boots are significantly faster (mean) than the combat boots. For Sara, the difference (of means) between the clubs is not large enough to achieve statistical significance. (Note: Because both studies are balanced and have nearly equal sample standard deviations, all the usual two sample t formulas give essentially the same quantitative answer.)

Below I will list 11 more student projects. Because of space limitations, I will not discuss the results of the data analysis.

14. **Maria** was a member of the UW's varsity soccer team. Her teams had two games per week, every Friday evening and Sunday morning. Maria wondered whether she played better in the evening or morning. As an indirect way of measuring this, she performed a study to see whether her ability to juggle a soccer ball was influenced by the time of day.
15. **Jennifer E.** was a softball player and wanted to investigate whether the type of bat, aluminum or wood, influenced how far she could hit a pitched ball.
16. **Doug** performed a study to investigate whether the type of dart he used, "bar darts" or his personal set, influenced how many rounds he required to complete the game "301."
17. **Kymn** was a member of the women's varsity crew and frequently worked out on a rowing simulation machine called an ergometer. She wanted to investigate whether the setting of the machine, there are two, influenced the time required to row a simulated 2,000 meters. No doubt because of her superb conditioning, Kymn's times on each setting showed little variation.
18. **Mei Lan, Sin Fai, and Todd** performed a study to investigate whether having a car's windows open or closed influenced the time required to accelerate from 40 to 65 miles per hour.
19. **Damion** performed a study to determine whether his smoking a cigarette or not influenced how long it took him to climb Bascom Hill on the Madison campus.
20. **Lisa J.**, a former horticulturist in my class, performed a study to investigate the effect of temperature on the growth of *Rhizoctonia solani* fungus.
21. **Jennifer C.** swam competitively in college and was a high school swimming coach when she enrolled in my course. She performed a study to compare two methods of starting a freestyle swimming race. She found that, for her, the traditional start was superior to the new "track start."
22. **Eric** was a punter on the varsity football team and performed a project to estimate how much farther he would punt the ball with two steps (before kicking) rather than one step.

Most projects with a numerical response are concerned with the center of the distribution; either larger or smaller is better. The next two projects provide examples in which spread is more important than center. The first project is an example that deals with variation from a fixed target and the second deals with the more common statistical problem of variation without a fixed target.

23. **Paul and Leslie** enjoyed playing the dart game Cricket, with each claiming to be the superior player. Hitting the twenty wedge on the dart board is particularly important for a successful game of Cricket. Each person attempted 78 throws, with each toss aimed at the 20 wedge. A response of 0 was noted if the throw hit the target wedge of 20. If the throw landed one wedge to the right of 20, the response was +1; if it landed two wedges to the right of 20, the response was +2, and so on. Darts that landed to the left of 20 gave responses of -1, -2, and so on, in the analogous manner. These data contained a wealth of information. Viewed as a dichotomy, Paul was better at hitting the 20, by a score of 36 to 10. (This pretty much settled who was the better player.) Paul's mean response was 0.31 and Leslie's was 0.65, indicating that both had a tendency to shoot to the right. Finally, Paul's mean absolute response was 0.62 and Leslie's was 1.71, indicating a substantial difference in variation.

24. **Kim** wanted to compare the distributions of the distance obtained when hitting a golf ball with two seven-irons; one made of steel and one of graphite. In golf, consistency with a seven-iron is more important than distance; if you want to hit the ball farther, use a six-iron. By examining dot plots, histograms, and various measures of spread, Kim concluded that there was no substantial difference in the spreads of the two distributions.

Most students have no trouble selecting appropriate trials for their projects, but a significant minority make a particular error, namely they make the trials “too small.” This error is particularly common for students studying darts or bowling. For example, a bowler might want to know whether the weight of the ball influences performance. What should the trial be: a frame or a game? Unfortunately, some students seem to reason as follows. More data gives us better answers, and defining a trial as a frame will yield more data than defining a trial to be a game. Therefore, I will have more data if the trial is a frame. Let us ignore the important issue that for trial equal to frame, there is no sensible way to measure the response—a ten and an eight are clearly much better than two nines, and a ten on one ball, a strike, is better than a spare. This is where randomization can help the student to see his or her error. If the trial is a frame, then when one randomizes it is likely that the bowler frequently will be forced to switch from one ball to the other. But no serious bowler would do that! (Similarly, a dart player would not change darts after every throw.) My advice to students is to let the goal of the study determine the size of the trial and be careful not to make the trial too small. For example, a bowler is interested in the score in a game of bowling, so any trial smaller than a game would be wrong. (In fact, a league bowler might be interested in the score for three games and choose three games to be the unit.)

5 Examining Assumptions

See [6] for a more detailed discussion of the ideas in this section.

As discussed in Section 2, for population inference to be valid, one needs to assume that the units are obtained by random sampling. If the units are trials then random sampling means that the trials are independent and identically distributed (i.i.d.). For a dichotomous response, i.i.d. trials are called Bernoulli trials. In this section I discuss student projects that examine whether a sequence of dichotomous responses can be viewed as Bernoulli trials. If the trials are Bernoulli trials, then population inference is valid. Note that, unlike earlier projects that compared data from two groups, the studies of this section have data from only one source.

The idea of a mathematical model for the outcomes of trials is difficult for many of my students to understand. For example, if I tell them that a random sample of 100 persons from a population yields 70 successes, they understand that $\hat{p} = 0.70$ might not be the same number as p , the population proportion. If, however, I say that a basketball player’s successive free throws are Bernoulli trials and that she makes 70 out of 100 shots, many of my students conclude that $p = 0.70$. It is strange; if $n = 1$ students readily understand that p need not be 1 or 0 (the two possible values of \hat{p}), but somehow as n grows the distinction between \hat{p} and p disappears, faster and more completely than the law of large numbers indicates.

Even though it is difficult to teach students a better understanding of Bernoulli trials, I believe that it is an important topic. Many sciences rely critically on mathematical models and I believe there is value in helping students obtain a better understanding of such models, even in the simple form of Bernoulli trials.

Below are some examples of sequences of dichotomous trials observed or obtained by my students.

25. **Mieke** defined a trial to be her playing a B-flat on her clarinet into a tuner. The tuner classified the note as sharp, flat, or perfectly in tune. A perfectly in tune note denoted a successful trial, while either of the other two classifications was labeled a failure.

26. **Mary** defined a trial to be her father’s morning wait for the bus. If the bus arrived within two minutes of the scheduled time, the trial was a success; otherwise, it was a failure.
27. **Matthew** defined a trial to be a live birth at a local hospital; female babies were successes and males were failures.
28. **Mark** and **Justin** defined a trial to be a bicyclist riding through a busy intersection in Madison. A rider with a helmet was a success and a rider without a helmet was a failure.
29. **Kristin** defined a trial to be rolling two small balls across the floor in her living room. A trial was a success if her dog, Muffin, fetched the red ball and a failure if she fetched the blue ball.
30. There have been numerous projects in which the trials are golf putts, basketball free throws, dart tosses, and other game activities.

I suggest two simple-minded ways to investigate the validity of the two assumptions of Bernoulli trials. First, compare the proportion of successes in the first and second halves of the study. Second, look at the 2×2 table that relates the current outcome to the previous outcome. The first comparison gives some insight into whether p remains constant and the second provides some insight into whether the trials have memory.

The dichotomous trials projects have been quite successful. The only serious problem was that, initially, my students chose really boring trials—tossing a coin, casting a die, throwing paper wads in a basket. It turned out that they had misinterpreted my instructions. They thought that I wanted them to be sure they had Bernoulli trials. Quite the contrary, often it can be very interesting to discover memory or a time trend in a process.

6 Conclusion

Confucius taught,

I hear, I forget.

I read, I remember.

I do, I understand.

I encourage you to include more “doing” in your introductory statistics classes. For me, assigning small student projects has been a very successful and rewarding way to incorporate more active learning into an introductory statistics course. An added bonus is that students get additional practice at writing.

If you decide to include small student projects in your course, I hope that the examples and comments in this paper will prove useful.

References

- [1] Ramsey, F.L. and Schafer, D.W.: *The Statistical Sleuth: A Course in Methods of Data Analysis*, Duxbury Press.
- [2] Rossman, A.: Book Review of *Statistics: Learning in the Presence of Variation*, *The American Statistician*, May, 1995, Volume 49, Number 2, pages 237–8.
- [3] Wardrop, R.: “A Radically Different Approach to Introductory Statistics,” University of Wisconsin–Madison, Department of Statistics Technical Report No. 889, 1992.

- [4] Wardrop, R.: “Guest Editorial: A New Approach to Introductory Statistics,” *The Journal of Undergraduate Math and Its Applications*, Spring 1995, Volume 16, No. 1, pages 1–8.
- [5] Wardrop, R.: “Student Sports Projects in a Statistics Course,” An invited paper presented at the Joint Statistical Meetings, Section on Statistics in Sports, 1996.
- [6] Wardrop, R.: “Bernoulli Trials: Do They Exist?” A contributed paper presented at the Joint Statistical Meetings, Section on Statistical Education, 1996.