BEARS IN SPACE: ACTIVITIES TO INTRODUCE BASIC IDEAS OF DESIGN

George W. Cobb, Mt. Holyoke College, USA
Weiwen Miao, Colby College, USA

Two elements of the reform movement in undergraduate statistics education are greater emphasis on data production (a curricular change) and greater emphasis on classroom activities (a pedagogical change). This paper presents a sequence of closely related activities, using simple equipment, for introducing students to the basic concepts of experimental design. The activities take as their nominal goal to understand the factors that influence the distance traveled by soft candy bears catapulted from a homemade launcher. They introduce students to variability and confounding, the need for a protocol, for replication and randomization, blocking as a strategy for reducing variability, and factorial crossing as a strategy for studying interaction. The analysis can be kept simple and informal, but the data also lend themselves to a more ambitious analysis, including, for example, both formal ANOVA and related diagnostics; in particular, the variation in launch distances is proportional to the mean distance, suggesting a transformation to logarithms.

INTRODUCTION

Using classroom activities stimulates students' desire to study statistics, which is one of the most important factors for success in learning. Asking students to collect data themselves not only enhances their interest in analyzing the data, but also emphasizes the importance of design. It convinces students that statistics is interesting and useful. It also illustrates how to apply statistical principles to solve problems in everyday life and encourages students to do so. When done in groups, activities help create an atmosphere of cooperation that leads to more active class discussions. In this paper, we present a simple inexpensive classroom activity, Bears in Space, to introduce some basic ideas of experimental design. For clarity, we have organized our exposition into a sequence of three activities and eight different topics for class discussion, but in practice there is considerable flexibility in the structure of the activities, and class discussions are never so neat as we pretend here.

PREPARATION

The activities are based on launching soft candy bears (sold in the US as "gummy bears") from a ramp in order to study a variety of factors that may affect launch distance. Launchers are made from popsicle or craft sticks and rubber bands (Figure 1). First, wind a rubber band enough times around one end of a stick so that the rubber band stays firmly

in place. Then, place that stick and another one together and wind a second rubber band tightly around the other end of the two sticks to bind them firmly together. Insert a thin pencil or small wooden dowel between the two sticks as a fulcrum.

Figure 1. Building a launcher

Put the launcher on a board or a rigid meter stick with blocks or books underneath one end (Figure 2). Load the gummy bear on the launcher, then launch. After each launch, record the distance, measuring from the *front* of the ramp, only in the direction *parallel* to the ramp.

Launch ramp

Books or blocks   Board or meter stick   Measure launch distance
from front of ramp
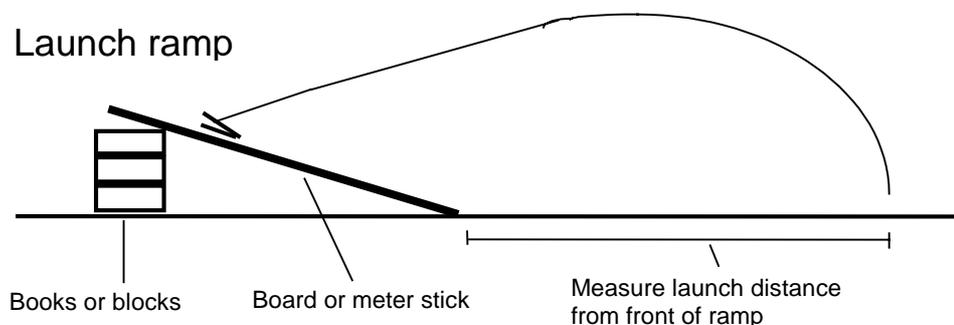
Figure 2. Setting up the launch

ACTIVITY 1

This experiment studies two possible factors that may affect launch distance: the launch angle (or number of books under the ramp) and the color of the gummy bears. First, divide the class into launch teams. (A team can be anywhere from 2 to 6 students.) Then, organize the teams; assign each student to one or more of the following jobs: hold the launcher on the ramp, load the gummy bear, launch, measure, and record the data.

Use a random number table or a coin toss to assign each team to produce data for one of the two launch angles: steeper ramp (four books under the ramp) and flatter ramp (one book under the ramp). Each team should do 10 launches: the first 5 with red bears, the last 5 with green bears. (Advise students that when launching the bears, they should use a coin instead of a fingertip to avoid ending up with sore fingers.) Usually it takes about 20 to 30 minutes for this activity, leaving enough time for a substantive discussion.

Discussion 1. The structural elements of an experiment

The response is launch distance; there are two conditions of interest: launch angle and bear color. Launch angle is an experimental condition with two randomly assigned treatments: one or four books under the ramp; bear color is a quasi-observational condition with two levels: red or green bears. Because the treatment (launch angle) is assigned to each team, the experimental unit here is a set of ten launches by a team, not a single launch.

Discussion 2. Confounding, randomization, and diagnostic time plots

After the activity, we ask each team to put its 10 distances and the average distances for red and green bears on the board. Usually the data indicate that green bears travel farther. Is the extra distance due to bear color or to the fact that each team launches red bears first? Generally, students realize that distances improve with practice, and time plots for the bear-launch data typically confirm this: later launches tend to be longer, inviting the conjecture that color and launch order are confounded. To protect against confounding and make causal inference possible, we need to randomly assign the conditions that we want to compare. As an improvement of the experiment, the order of launches within each team should be randomized.

Discussion 3. Getting the experimental units right

Next, ask students to forget the bear color and suggest ways to judge the effect of launch angle. Generally they come up with two methods: (1) compare individual launches, while ignoring team differences; and (2) compare team summaries, while ignoring individual launches. Deciding between the two strategies reinforces the importance of recognizing the experimental unit. (Here the unit is the team, and we need the size of variation between units (teams) in order to judge whether launch angle has an

effect. Moreover, in this experiment, the number of teams matters more than the number of launches per team, since the amount of information we get from an experiment depends on the number of units. The first analysis, of individual launches, gives up the essential information about the variation between teams and hence is seriously flawed.)

Discussion 4. Protocol

In order to judge the effects of the treatment (launch angle), it is important to keep all other possible influences on the response as nearly constant as possible. (That's why, for example, each team should let the same person do all the launches.) One important way to control all other possible influences on response is to develop a careful *protocol* that tells exactly how the experiment is to be done. To write a good protocol, students should first *identify the important sources of variation.* Here, these include the force of the launch, position of the fulcrum, effect of practice, position of the bear on the launcher, how the launcher is held, etc. After identifying the sources,
each team should write out an experimental protocol designed to keep launch conditions as nearly constant as possible.

Discussion 5. Blocking

Using teams introduces two kinds of variation in the bear-launch experiment: within-team variation and between-team variation. Recording the team summaries in a side-by-side dot plot usually indicates that the variation between teams is large. Even though assigning launch angles randomly to teams gives protection against confounding, the variation between teams is so large that it would have taken a large number of units (teams) to be confident of detecting differences due to launch angle. In practice, though, there are limits on the number of units. (Here, for example, the number of teams can be at most the size of the class.) One solution to this problem is to design a study to translate the launch angle, which is the between-team factor in the previous experiment, to a within-team factor, by blocking. In the new design (Activity 2), each team (block) is reused to provide two sets of five launches (units) . This makes it much easier to detect the effects of launch angle *without increasing the number of teams.* Students should notice (1) that differences between teams are not part of the research question; rather, those differences are a nuisance, because they make it harder to answer the question of

interest, and (2) that, as a rule, blocks in a block design come from some such nuisance influence.

ACTIVITY 2

Each team does two sets of five launches, one set with four books under the ramp, and the other with just one book. Each team should randomize the order of the treatments, following the protocol developed before to carry out the launches. Table 1 gives data from a class with 6 teams.

	average distance (inches)					
Team	1	2	3	4	5	6
1 book	87.0	43.0	87.0	38.0	49.0	44.0
4 books	246.0	67.0	244.0	81.0	103.0	64.0
Difference	159.0	24.0	157.0	42.0	54.0	20.0

Table 1.   Sample data for complete block design

From Table 1, it seems reasonable to conclude that launch angle makes a difference, as a formal ANOVA confirms.

Discussion 6. Factorial crossing and interaction

At this point, the students are familiar with bear-launches, but not yet bored with bears, so one can use the activity to introduce the two-way factorial design. We ask students to think about the effect of the position of the launcher on the ramp: do they expect to gain more distance by launching from the back of the ramp? Notice that when launched from the back of the ramp, the bears start higher up, and so would be likely to travel farther before landing. On the other hand, the rules say you have to measure distances from the front of the ramp, so there is a cost to moving back in order to gain height. Furthermore, there is the possibility of interaction. In the previous experiment, it was natural to think that launch angle and bear color don't interact. What will happen if we study launch angle and the position of the launcher on the ramp? Will the extra distance (if any) be the same whether we have one or four books under the back of the ramp? These questions lead to a two-way factorial design. For example, if we use 3 levels for launch angle and 2 levels for launcher position, then *crossing* the two factors of

interest (using all possible combinations of levels) gives 6 combinations of the launch angle and launcher position.

ACTIVITY 3

Each team carries out 6 sets of 4 launches each, one set for each of the 6 combinations of launch angle and position of the launcher. Launches should follow the protocol from before, and the order of the 6 conditions should be randomized. Table 2 shows averages of 4 launches for all 6 conditions from an actual class; Figure 3 shows the corresponding interaction graph.

Position	Number of Books		
	0	4	8
Front	44.9	133.4	126.8
Back	29.0	115.4	140.5
Diff	-15.9	-18.0	13.7

Table 2.   Sample data for two-way
          factorial design

figure 3   Interaction graph
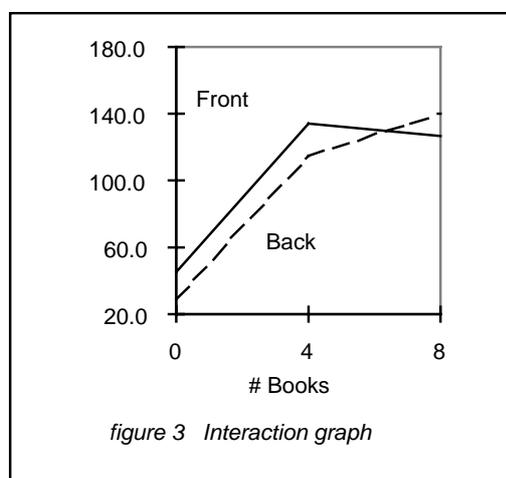
As the differences in Table 2 reveal, the distance one gains or loses by launching from the back of the ramp instead of the front is different for the different launch angles. This fact helps students understand that interaction is actually a "difference of differences." No interaction means that the differences are the same for all the launch angles; in the interaction graph, the lines should be parallel. A formal ANOVA confirms the presence of interaction.

Discussion 7.   Extensions

Other factors may also affect launch distance, for example fulcrum size and fulcrum position. The fulcrum (the wooden dowel inserted between two sticks) can be small or large, and it can be put far away from the gummy bear, in the middle of the

launcher, or close to the bear. Students can design an experiment to test for interaction between the position of the fulcrum and the launch angle; or as a small project students might design a three-factor experiment to find the combination of launch angle, launcher position, and fulcrum position that together produce the longest launches.

Discussion 8. Diagnostics, and transforming to equalize spreads

The analysis of the gummy bear data can be kept simple and informal, but instructors can also use the data to introduce formal ANOVA and related diagnostics. In particular, the gummy bear data usually indicate that the standard deviation (SD) for each team is different. Typically, large means go with large SDs, small with small. This suggests that transforming to logarithms will equalize the SDs. In this context, it is natural to introduce the power family of transformations and diagnostic plot for choosing the power.

CONCLUSION

In our experience, the combination of activities described here is an effective vehicle for getting students involved in discussions that bring out the basic concepts of experimental design. They find the activities engaging, are eager to discover the answers, and in the process find the design concepts relevant and accessible. The activity is simple and flexible enough to be easily adapted to a variety of teaching situations.

REFERENCE

Cobb, G. W. (1998). *Introduction to Design and Analysis of Experiments*. New York: Springer-Verlag.