Principles of Experimental Design

Bret Hanlon and Bret Larget

Department of Statistics
University of Wisconsin—Madison

November 15, 2011

Experimental Design

- Many interesting questions in biology involve relationships between *response variables* and one or more *explanatory variables*.
- Biology is complex, and typically, many potential variables, both those measured and included in an analysis and those not measured, may influence the response variable of interest.
- A statistical analysis may reveal *an association* between an explanatory variable and the response variable.
- It is very difficult to attribute causal effects to observational variables, because the true causal influence may affect both the response and explanatory variable.
- However, properly designed experiments *can reveal causes of statistical associations*.
- The key idea is to reduce the potential effects of other variables by designing methods to gather data that *reduce bias and sampling variation*.

Case Studies

We will introduce aspects of experimental design on the basis of these case studies:

- An education example;
- An *Arabidopsis* fruit length example;
- A starling song length example;
- A dairy cow nutrition study;
- A weight loss study.

Biology Education

Case Study

- A researcher interested in biology education considers two different curricula for high school biology.
- Students in one school follow a standard curriculum with lectures and assignments all from a textbook.
- Students in a second school have the same lectures and assignments, but spend one day each week participating in small groups in an inquiry-based research activity.
- Students from both schools are given the same exam.
- Students with the standard curriculum score an average of 81.2 and the group of students with the extra research score an average of 88.6; hypothesis tests (both a permutation test and a two-independent-sample t-test) have very small p-values indicating higher mean scores for the extra research group.

Arabidopsis Example

Case Study

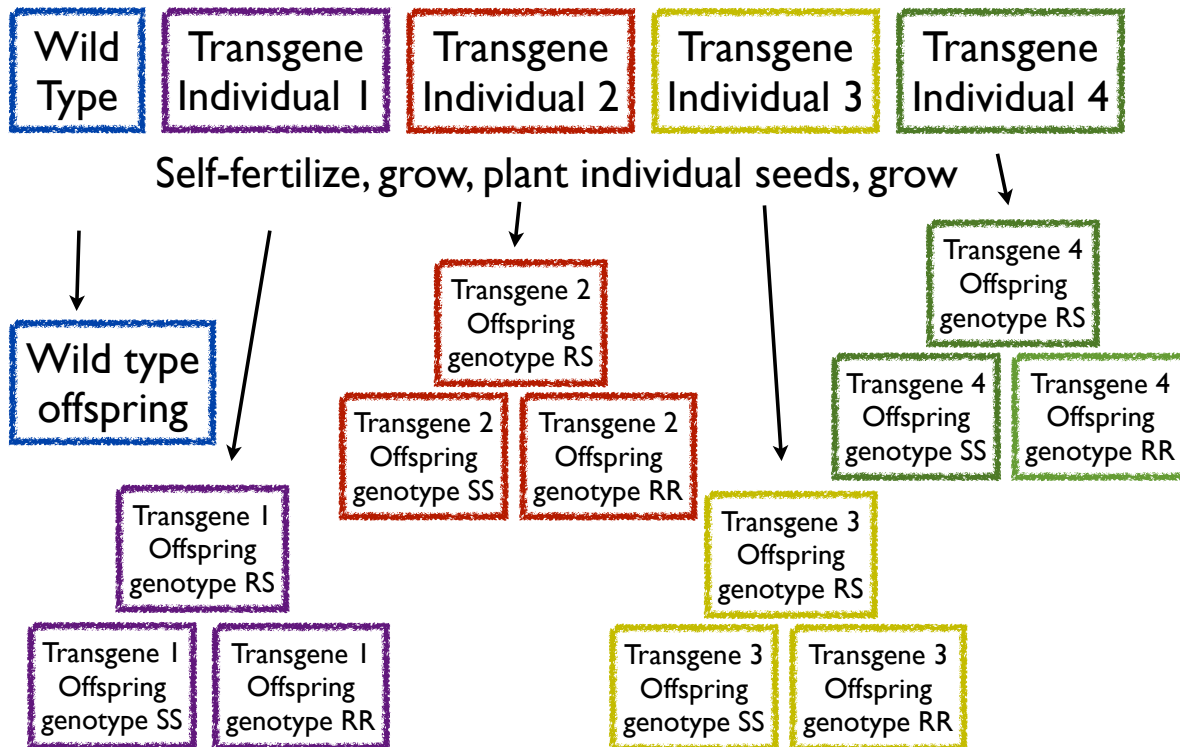A researcher conducts an experiment on the plant *Arabidopsis thaliana* that examines fruit length.

- a gene from a related plant is introduced into the genomes of four separate *Arabidopsis* plants;
- each of these plants is the progenitor of a *transgenetic line*.
- an additional *Arabidopsis* plant is included in the experiment, but does not have the trans-gene introduced;
- these five plants represent the *T1* generation;
- each T1 plant is grown, self-fertilized, and seed is collected;
- a sample of 25 seeds from each plant are potted individually, grown, and self-fertilized;
- these plants are the *T2* generation;
- the length of a sample of ten fruit is measured for each T2 plant.

Arabidopsis Genotypes

Case Study

- The gene (call it R) is inserted into one locus for each T1 transgenic plant;
- The sister chromosome will not have the transgenic insertion;
- After self-fertilization, we would expect:
  - about 25% of the T2 plants to have 0 copies of the transgenic element; — this is genotype SS
  - about 50% of the T2 plants to have 1 copy of the transgenic element; — this is genotype RS
  - about 25% of the T2 plants to have 2 copies of the transgenic element; — this is genotype RR
- The genotype of each T2 plant is inferred by collecting and growing a sample of its seeds.
- All wild type offspring have geneotype SS.

Arabidopsis Experiment

Wild Type | Transgene Individual 1 | Transgene Individual 2 | Transgene Individual 3 | Transgene Individual 4

Self-fertilize, grow, plant individual seeds, grow

Wild type offspring

Transgene 1 Offspring genotype RS

Transgene 1 Offspring genotype SS | Transgene 1 Offspring genotype RR

Transgene 2 Offspring genotype RS

Transgene 2 Offspring genotype SS | Transgene 2 Offspring genotype RR

Transgene 3 Offspring genotype RS

Transgene 3 Offspring genotype SS | Transgene 3 Offspring genotype RR

Transgene 4 Offspring genotype RS

Transgene 4 Offspring genotype SS | Transgene 4 Offspring genotype RR

Starling Song

Case Study

- Starlings are songbirds common in Wisconsin and elsewhere in the United States.
- Male starlings sing in the spring from a nest area when they attempt both to attract females as potential mates and to keep other males away.
- Male starlings sing in the fall when they are in flocks of other male birds.
- It is difficult to categorize a single song as "spring-like" or "fall-like", but characteristics of song can be different at the two times.
- One simple song characteristic is the length of the song.
- In an experiment, a researcher randomly assigned 24 starlings into two groups of 12.

Starling Song (cont.)

Case Study

- All measurements are taken in animal observation rooms in a research laboratory.
- The spring group was kept in a *spring-like environment* with more light, a nest box, and a nearby female starling.
- The male group was kept in a *fall-like environment* with less light, no nest boxes, and in the proximity of other male birds.
- Each bird was observed and recorded for ten hours: birds sang different numbers of songs, and the length of each song was determined.
- Each bird sang from between 5 and 60 songs.
- (In the actual study, characteristics of the songs beyond their length were of greater importance.)

Dairy Cattle Diet Example

Case Study

- In a study of dairy cow nutrition, researchers have access to 20 dairy cows in a research herd.
- Researchers are interested in comparing a standard diet with three other diets, each with varying amounts of alfalfa and corn.
- In the experiment, the cows are randomly assigned to four groups of 5 cows each;
- Each group of cows receives each of the four diet treatments for a period of three weeks; no measurements are taken the first week so the cow can adjust to the new diet.
- The diets are rotated according to a Latin Square design so that each group has a different diet at the same time.
- Response variables include milk yield and abundance of nitrogen in the manure.

Latin Square Design

> **Definition**
>
> A *Latin square design* is a design in which three explanatory variables (typically one treatment and two blocking), each of which is categorical with the same number of levels (in this example, four), so that each pair of variables has the same number of observations for each possible pair of levels. Treatments are placed in a square so that each row and column contains each treatment once.

Diets are named A, B, C, and D. Each group of cows gets all four diets, but in different orders.

		Time Period		
Group	First	Second	Third	Fourth
1	A	B	C	D
2	C	A	D	B
3	B	D	A	C
4	D	C	B	A

Weight-loss Study

> **Case Study**
>
> - Researchers in the Department of Nutrition recruited 60 overweight volunteers to participate in a weight loss study.
> - Volunteers were randomly divided into two treatment groups.
> - All subjects received educational information about diet.
>   - one treatment group was instructed to count and record servings of each of several food types each day;
>   - the other treatment group was instructed to count and record calories consumed each day.
> - Subjects were not aware of the instructions given to members of the other group.

Confounding

Definition

A *confounding variable* is a variable that masks or distorts the relationship between measured variables in a study or experiment. Two variables are said to be *confounded* if their effects on a response variable cannot be distinguished or separated.

Problem

1. What are possible confounding variables that may explain the differences in test scores in the education example?
2. What potential confounding factors are researchers trying to avoid with the Latin square design for the dairy cow nutrition study?
3. What are potential confounding factors in the weight loss example?

Experimental Artifacts

Definition

A *experimental artifact* is an aspect of the experiment itself that biases measurements.

Example

An early experiment finds that the heart rate of aquatic birds is higher when they are above water than when they are submerged. Researchers attribute this as a physiological response to conserve oxygen. In the experiment, birds are forcefully submerged to have their heart rate measured. A later experiment uses technology that measures heart rate when birds voluntarily submerge, and finds no difference in heart rates between submerged and above water groups. This suggests that the stress induced by forceful submersion rather than submersion itself caused the lowering of heart rate in the birds.

Experimental Artifacts

Problem

1. What potential experimental artifacts might be present for the starling song experiment?

2. In designing an experiment to study the natural behavior of living organisms, what trade offs are there between gathering data in nature or in a laboratory setting?

Control Groups

Definition

A *control group* is a group of individuals that do not receive the treatment of interest, but otherwise experience similar conditions as other individuals in the experiment or study.

Problem

1. What is the control group in the arabidopsis experiment?

2. Which comparison between the control group and another group may be most informative about the effects of the experiment?

3. Is there a control group in the education example? Discuss.

4. Is there a control group in the starling song example? Discuss.

5. Is there a control group in the dairy cow nutrition example? Discuss.

Randomization

Definition

Randomization is the random assignment of individuals to different treatment groups.

- The purpose of randomization is so that the effects of potential confounding variables, whether these variables are known or not, are likely to be divided fairly evenly across treatment groups.
- Of course, for any particular variable, the values for individuals in each sample will not be exactly balanced for each specific assignment to treatment groups.

Stratified Randomization

- Treatment groups can be partially randomized; for example, if in the dairy cow example it was known that there were 8 cows in their first milking and 12 cows not in the first milking, the 8 primiparous cows could be randomly assigned to two to each group and the 12 multiparous cows could be randomly assigned three to each group.
- This is an example of a *stratified random sample*.
- Stratification may be warranted if a variable is known to affect the response variable of interest in order to lessen the amount of confounding that might be caused by an unlucky complete randomization.
- The trade-off is that other variables may be *more likely* to be unbalanced in the assigned treatment groups.

Randomization Examples

Problem

- Discuss the role of randomization in each of these examples:
  1. Education example;
  2. Starling example;
  3. Dairy cow example;
  4. Weight-loss example.
- When is randomization practical?
- How might stratification have been done in each case?
- What are the trade offs?

Blinding

Definition

An experiment is *blinded* if the subjects do not know which treatment group they are in. An experiment is *double blind* if both the subjects and the researchers measuring responses are unaware of the treatment group for each subject.

- Blinding is meant to protect against experimental artifacts that could be caused by knowledge of the subjects (or the researchers doing the study who may be subconsciously influenced to expect to see something consistent with a hypothesis).

Blinding (cont.)

> **Problem**
> Discuss the application (actual or possible) of blinding in each example:
> 1. Education example;
> 2. Starling example;
> 3. *Arabidopsis example*;
> 4. Dairy example;
> 5. Weight loss example.

Replication

> **Definition**
> *Replication* is the repetition of each treatment on multiple independent experimental units.

- Replication is necessary for statistical inference, because it allows for variation between treatment groups (the variation of interest) to be compared to variation within groups.
- Inference for a mean requires the ability to estimate the size of a typical deviation (σ) and the mean (μ). If there was only one replicate, then we could estimate the mean, but not variation around the mean.
- Beware of *pseudoreplication*!

Pseudoreplication

> **Definition**
>
> *Pseudoreplication* (see page 97, Interleaf 2) is when individual measurements are not independent, but are treated as if they are.

- In the bird song example, there are many songs, but the songs are not independent, as songs from the same bird may be more similar than songs from different birds.
- In the dairy cow example, their might be daily measurements, but the sampling unit is the cow.
- In the *Arabidopsis* example, there are 1250 different fruits, but they are not independent as they are grouped according to the T2 plant from which they come.
- In the education example, the "treatment" is applied to the school, so there is no replication at all, with only one school per treatment group.

Balance

> **Definition**
>
> A design is *balanced* when each treatment group has the same size.

- When population standard deviations are equal, the standard error for the difference is smallest when both sample sizes are equal. (When population standard deviations are not equal, this is not necessarily true—with unequal standard deviations, it may be beneficial to sample more individuals from groups with higher variance).
- Data can be analyzed if sample sizes are balanced or not, but some methods are more accurate when samples are balanced.
- In the cow nutrition example, balance is forced because each animal receives each treatment and is measured during each time period.

Blocking

Definition

Blocking is placing sampling units into groups that are similar with respect to one or more covariates. Treatments are assigned at random within the blocks.

- The paired design is an extreme form of blocking where each pair of measurements form a block of size two.
- Blocking on the basis of one factor assures that the one factor is close to balanced in each treatment group.
- If you attempt to block on multiple factors, the number of blocks grows large and there are insufficient individuals to place into each block.
- Blocking is an attempt to directly control for the effects of a factor.
- Blocking and randomization are two methods to reduce bias from confounding factors, but they are in tension with one another; the more blocking that is employed, the less remaining allocation to groups is relegated to randomization.

Examples

Problem

1. In the dairy cow example, each cow is measured for each of four treatments. Explain how this is related to blocking.
2. In agricultural and ecological experiments, it is common to select blocks as regions where conditions are similar, and then split the blocks into smaller plots or subplots which are then placed into treatment groups.
3. The *Arabidopsis* experiment resulted in 1250 fruit lengths. How are these measurements blocked?
4. The starling birdsong experiment resulted in hundreds of different songs. How are these songs blocked?

Factors

Definition

A *factor* is a single categorical variable. The possible values of a factor are called *levels*.

- Some studies involve multiple factors.
- In the dairy cow study, the factor of interest is diet (with four levels, A, B, C, and D), but there are also two *nuisance factors* (cow group and time period) that are not of direct interest, but the effects of which need to be controlled.

Definition

In a *complete factorial design*, there are individuals for all treatment combinations.

Problem

1. The Latin square for the cow example is not a complete factorial design. Explain why a complete factorial design cannot be used in this setting.

Interaction

Definition

Two (or more) explanatory variables have an *interaction* if the effect one has on a response variable depends on the value of the other variable (or variables).

- It is necessary to make observations at all treatment combinations of the corresponding variables to estimate interactions.
- *Interaction plots* are useful informal graphical devices for exploring interactions in data.
- In an interaction plot with two explanatory factors, the levels of one factor appear on the x axis, the response is on the y axis, and the other explanatory factor is represented with color or a symbol of the plotted points. Lines often connect the means within each group. The further the lines are from being parallel, the stronger the indication of an important interaction.

Interaction Plots

- The following plot (data from Example 18-3) shows the response (area of red algae) in an experiment in the intertidal habitat of coastal Washington on the basis of two factors.
- Experimental location is either just above the low tidal zone or midway between the low and high tidal zones, and each location is either accessible to herbivores or not.
- The plot indicates that the herbivore treatment has little effect at mid tidal zones, but that excluding herbivores results in larger algal growth at low tidal zones.
- Formal statistical inferences depend on more than observing the lack of parallelism in the plotted lines.

Interaction Plot
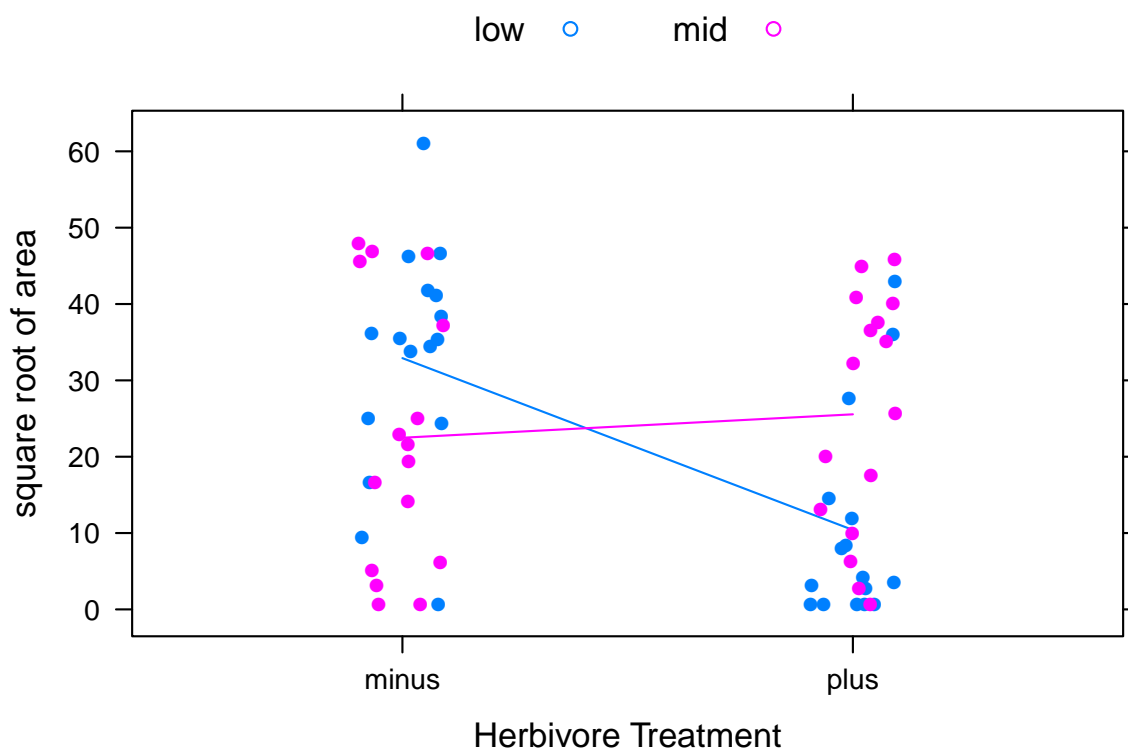
What you should know

You should know:

- how to define and recognize all of the concepts related to experimental design in these notes;
- how to identify from a description of an experiment possible confounding variables;
- how to distinguish between experimental and observational variables;
- how to recognize pseudoreplication;
- how to interpret interaction plots;
- when it is appropriate to make causal inferences.