AN INTRODUCTION TO HYPOTHESIS-TESTING AND EXPERIMENTAL DESIGN

INTRODUCTION

Ecologists gain understanding of the natural world through rigorous application of the scientific method. As you may recall, the scientific method instructs researchers to 1) observe nature carefully, 2) ask an interesting question, and 3) formulate a hypothesis. The **hypothesis** is an explicit statement as to what we believe to be true about the observed phenomena. The scientific method then instructs researchers to test the hypothesis. Testing usually involves designing a method or protocol for collecting information (data) that will allow us to evaluate the hypothesis, and finally, to accept or reject it.

We generally form a hypothesis based upon observations of natural phenomena during field (or laboratory) studies. Often these observations include the detection of distinctive patterns (or differences) in nature. For example, you might notice that the number of plants differs on north and south-facing slopes. You might also notice that each slope receives a different amount of sunlight during the course of a day. It is very important that the observed pattern be real and worthy of study. Therefore, ecologists are careful to measure or **quantify** their observations. In our example, you might count the number of trees along three 20-m paths (transects) on each slope. You could also quantify the number of shrubs in three 1-m^2 areas (quadrates) and measure the amount of sunlight reaching the ground using a light meter.

In order to tell if the two slopes differ with respect to tree, shrubs or light, you must show that the differences you see are greater than would be expected by chance. This is the province of **statistic analysis** and **experimental design**.

THE NULL HYPOTHESIS AND HYPOTHESIS-TESTING

The way in which we determine if a subjective impression (“educated guess” or hypothesis) about an observation is correct is by performing a hypothesis test. This involves the formation of a **null hypothesis**, collection of unbiased observations, statistical testing, and acceptance or rejection of the null hypothesis. The null hypothesis (symbolically expressed as **Ho**) has the following form:

H_O: The difference between two or more data sets is no greater than that expected by chance.

A statistical test of the null hypothesis results in either its acceptance or rejection. Accepting the null hypothesis means that there is no statistical difference between the two data sets. Rejection means that the probability that chance is operating to give the observed difference is so small that we are willing to discount it, and we assume that the difference is real. Rejection of the null hypothesis allows one to accept the **alternative hypothesis**, which has the following form:

H_A: The difference between two or more sets of data is so great that it is unlikely to have occurred due to chance.

It is important to realize that the terms accept and reject are used instead of prove or disprove in relation to the null hypothesis. This is because we are dealing with mathematical probabilities, not certainties. *No matter how great the difference between two data sets, there is still a probability that the difference is due to chance*. In other words, if there is a 98% probability that there is a difference, then there is also a 2% probability that there is no difference. So, we cannot prove or disprove a hypothesis with 100% certainty but we can make the decision, based on an overwhelming probability, to accept or reject a hypothesis.

EXPERIMENTAL DESIGN

In biology, experiments may be designed to measure a pattern in the environment or to apply a specific treatment to carefully isolated experimental units. Experiments designed to measure patterns are termed "**mensurative**"; whereas experiments involving treatments are termed "**manipulative**". An **experimental unit** is that "thing" being observed or measured. It may be a plot in the forest, an aquarium full of fish, or a single human being. A **treatment** is the application of certain conditions (or factors) to the experimental unit. In manipulative experiments, the investigator keeps all external factors constant and manipulates only the desired treatment factor(s). In mensurative experiments, the investigator does not apply the treatment. He/she merely observes the results of a treatment previously applied by Nature or some human disturbance. In either type of experiment, the effects of the treatment on the experimental unit are determined by measuring **response variables**.

Consider an experiment testing the influence of sunlight on the growth of sugar maple (*Acer saccharum*) seedlings. An investigator establishes many 2x2 m plots in the forest and thins the canopy trees to expose some plots to low sunlight and others to high sunlight. In some plots, the canopy trees are not thinned at all. These plots are termed "**controls**" because no treatment was applied. Over time, the investigator measures the height of red pine seedlings in each plot. Each plot is an experimental unit; light intensity (or thinning) is the treatment; and change in height over time (growth) is the response variable. This experiment is manipulative because the investigator applied thinning to specific plots. If the investigator had merely gone out to a forest and selected plots that had varying densities of canopy trees (perhaps as a result of tree fall from heavy winds), the experiment would be considered mensurative.

Within any population, there exists a certain amount of difference from one individual object to the next. In the previous example, sugar maple seedlings in a given plot would have a variety of heights (or growth rates). This **variability** can be the result of many factors (ex: genetics, environmental influences, etc.). Our ability to distinguish between variability caused by a specific influence (i.e., our treatment) and variability caused by other sources is a fundamental requirement for a successful study. The sources of variability present in any study, other than the treatment, can make results difficult or impossible to interpret. The purpose of experimental design, therefore, is to minimize, or hold constant, all other sources of variability, *except for those that we wish to monitor*. The features of experimental design that address the problem of variability are **sample size, replication, control, randomization,** and **interspersion.**

SAMPLE SIZE

A collection of measurements taken from an experimental unit is termed a **sample**. This sample represents a subset of the population being studied. In order to estimate the variability within the population (or treatment), it is necessary to have samples from many experimental units. The total number of experimental units sampled is termed the **sample size**. The larger the sample size, the better one is able to account for variability within a population. In some ecological systems (ex: large lakes, forests, oceans), it is impossible to get samples from more than one experimental unit. The ability to sample multiple experimental units is limited by practical constraints (limited personnel, resources, time, or money). In these systems, variability must be estimated by taking samples over time and using specialized statistics.

REPLICATION

To detect differences between treatments, it is necessary to assess the variability within each treatment. The only way to do this is to have multiple experimental units for each treatment. The sample collected from each experimental unit within a treatment is termed a **replicate**. For the purposes of statistical analysis, *a minimum of three replicates* is necessary for each experimental treatment. In our previous example on the growth response of sugar maple seedlings to sunlight, the experimental units to be replicated would be the plots of seedlings exposed to each level of sunlight. It would be inappropriate to take several measurements of one experimental unit (e.g., one plot in low sunlight) and consider them as replicates of the entire treatment level. This is an example of **pseudoreplication**.

CONTROL

The control serves as a standard against which the experimental units are compared. Controls are experimental units that are identical to those receiving treatments except that they do not receive the **treatment factor**. In our sugar maple seedling example the amount of sunlight is the treatment, and therefore we would have one plot that receives no sunlight as a control. All other manipulations (i.e., watering, fertilization, etc.) should remain constant for all experimental units. Controls help to identify those changes that might occur as a result of factors that the experimenter could not hold constant, such as daily temperature fluctuations. Experimental units assigned to be controls are replicated just as experimental units receiving a treatment.

INTERSPERSION AND RANDOMIZATION

The assignment of treatments and controls to experimental units must be such that there is no interaction between experimental units and therefore, all units are likely, on average, to receive the same environmental conditions. This placement is referred to as **interspersion.** The way in which interspersion is achieved is through **randomization.** Ideally, the assignment of treatments and controls is done in a purely random fashion which allows chance to determine their placement and to determine which experimental units will receive which treatment. For large numbers of replicates this works well since it is not probable that similar treatments would all end up clumped together. For smaller numbers of replicates (as is often the case in ecological studies) a **systematic** alternating of treatments and controls is often necessary (see Figure 1.1). While desirable, interspersion is often not possible in mensurative experiments.

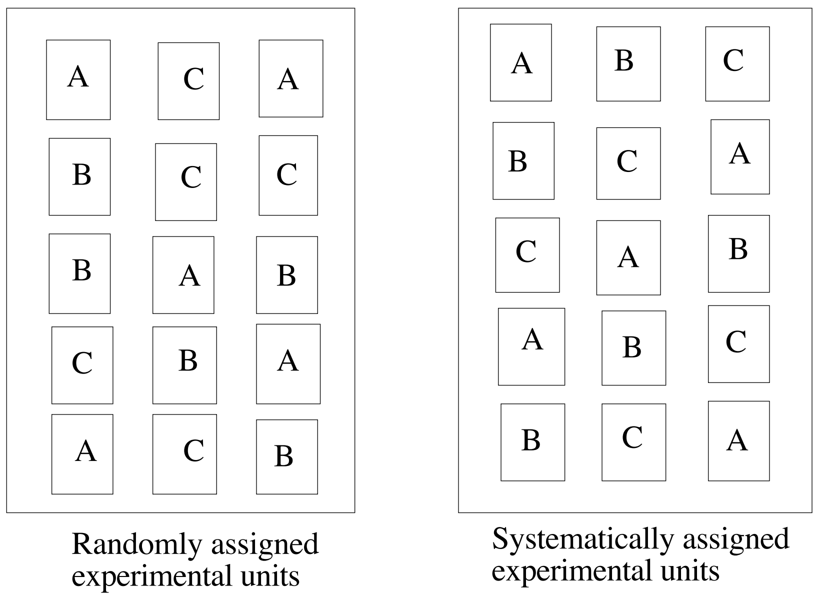


Figure 1.1

DISCUSSION QUESTIONS:

1. If one were planning to conduct a study investigating the effects of the salting of highways on stream "health", what might you measure? What would be your experimental unit?

2. Write a null and alternative hypothesis for the effect of highway salt on stream health.

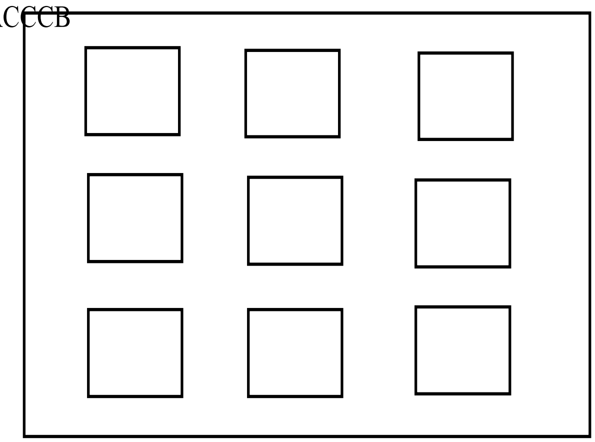
3. Identify the pseudoreplication in the situation given below:

   In studying the effects of sewage on alga growth in rivers an investigator sets up a site above and below a sewage treatment plant on a river. Several samples are taken upstream of the plant and several samples are taken downstream.

4. What are some sources of variability in the experiment on the effects of sunlight on the growth of sugar maple seedlings example given in the text?

5. You used a random method of interspersing your experimental units and came up with the following assignments:

AABBACCCB



Can you foresee any problem with this experimental design?

How might you solve this problem?