environment in which the experiment is run. These would include things like temperature, air pressure, and lighting conditions, but also things like the nutrition and habitats of living subjects, the timing of treatments, and other conditions that are external to the subjects themselves. Ideally, these environmental conditions will be the same for all subjects in the experiment; otherwise, we might not be able to eliminate the possibility that a difference in the results obtained from different subjects is due to differences in their environments. This is why experiments are often performed in labs, where the environment is carefully controlled and maintained to be as identical as possible for all treatments.

Individual subjects are not identical, so how do we determine whether a difference between the results from two subjects is not due to the fact that they are simply different individuals? Good experiments therefore need to control for individual variation. For instance, many experiments use model organisms, such as specially bred strains of mice or bacteria, where all individuals are as close to being genetically identical as they can be. Even if your mice are genetically identical, they might still be different due to the different lives that they have led. Thus, you would also want to make sure that your mice were of similar size, age, and health.

In many cases, you cannot simply select subjects that are nearly identical. This is especially true for clinical studies of human health. How does one control for individual variation then? In such cases, a researcher can design an experiment to minimize **bias**, or some kind of nonrandom variation in the treatments. For instance, if your control mice were all males, but your experimental mice were all females, that would be a very biased distribution of sexes in your treatments, and you would not be able to exclude the possibility that the differences in the results between your controls and experimentals were due to sex differences. In experiments like clinical trials, bias is avoided by **randomization**; subjects are assigned randomly to a treatment group, and with large enough sample sizes each treatment should have a set of individuals who, while not identical to each other, vary in roughly the same way from treatment group to treatment group.

Finally, good tests are or can be repeated, by the original researcher or by those in another lab. The terms **repetition** and **replication** are often used loosely for either kind of repeating an experiment, but typically repetition refers to essentially having multiple instances or runs of an experiment when it is carried out, whereas replication typically refers to someone else carrying out the same experiment to see if they get similar results. Note that, in this case, repetition might mean running an experiment several times in sequence, or having multiple control and treatment groups. In many cases, simply using multiple subjects is a form of repetition.

Activity

A) *For the three experiments in the previous activities, identify the additional information you would need if you were going to run these experiments yourself.*

B) *What would you need to do in each of the experiments to execute them and to control for confounding variables?*

C) *How many individuals would you need, and how would you take into account the effects of differences between individuals?*

A Closer Look: Descriptive Statistics and Calculating Means and Standard Deviations

An important part of experimental design is determining what you will measure and how you will measure it. After you make your measurements, how will you interpret them? For instance, in the plant experiment, when we measure the heights of our plants, it is very likely that they will not all be the same height, but how different should they be for us to say whether the prediction of a hypothesis is matched or not? If we use many plants in each treatment, we will need a way to describe the growth of each treatment as a whole. Descriptive statistics help us to do that. Two of the most basic descriptive statistics are the mean and the standard deviation. The **mean** is the average value of a measurement (say, of a particular trait) for all of the sampled individuals. While the mean is helpful, it doesn't provide a very complete picture of variation in this trait in the population. In other words, we might glean something from the fact that the mean is large or small, but is that mean derived from a sample that varies very little from the mean or that has a wide range of variation in this trait? A useful statistic for understanding the variation of a trait among sampled individuals around its mean is the **standard deviation**. The standard deviation is essentially the average amount by which individuals vary from the mean. If the range of variation is wide, then the standard deviation will be large relative to the mean; if there is little variation in the sample, then the standard deviation will be small relative to the mean.

The formulas below are for calculating means and standard deviations.

Mean: $\overline{X} = \dfrac{\sum x_i}{n}$ Standard deviation: $\sqrt{\dfrac{\sum (x_i - \overline{X})^2}{n-1}}$

The mean is fairly straightforward and probably familiar to most students, even if its representation in the above formula is not. The mean takes the sum of all the individual values for a given variable (in this case, x, with each individual's value for x given as x_i); that sum is Σx_i in the formula. We then divide that sum by the number of individuals in the sample, or n.

The formula for standard deviation shows some similarities to the formula for the mean: there is a sum being made that involves x_i, and we are dividing by a number close to the sample size ($n-1$). The main differences are that (a) we are subtracting the mean from x_i, (b) we're squaring that difference, and (c) we're taking the square root of the whole calculation. The first difference should make sense: we are interested in how individuals vary from the mean, on average, so we can start by calculating the difference between each individual's value for x and the mean. But why square this difference? The reason is that individuals vary both by being less than the mean and by being greater than the mean. Thus, if we just take differences from the mean, we will end up with some positive numbers and some negative numbers, and the sum of those will be zero. By squaring the difference, we eliminate the sign, and all squared differences will be positive. Thus, the resulting sum will be positive. By taking the square root, we are essentially "undoing" that squaring and ending up with a number that reflects how the population tends to vary from the mean.

The standard deviation is useful for describing how much a sample varies. Here is a simple example to illustrate this.