Chapter 13 Summary
Experiments and Observational Studies

What have we learned?

- We can recognize sample surveys, observational studies, and randomized comparative experiments.
    - These methods collect data in different ways and lead us to different conclusions.
- We can identify retrospective and prospective observational studies and understand the advantages and disadvantages of each.
- Only well-designed experiments can allow us to reach cause-and-effect conclusions.
    - We manipulate levels of treatments to see if the factor we have identified produces changes in our response variable.
- We know the principles of experimental design:
    - Identify as many other sources of variability as possible so we can be sure that the variation in the response variable can be attributed to our factor.
    - Control the sources of variability we can, and consider blocking to reduce variability from sources we recognize but cannot control.
    - Try to equalize the many possible sources of variability that cannot be identified by randomly assigning experimental units to treatments.
    - Replicate the experiment on as many subjects as possible.
- We've learned the value of having a control group and of using blinding and placebo controls.
- We can recognize problems posed by confounding variables in experiments and lurking variables in observational studies.

Observational Studies

- In an observational study, researchers don't *assign* choices; they simply observe them.
    - The text's example looked at a student of the relationship between music education and grades.
    - Since the researchers did not assign students to get music education and simply observed students "in the wild," it was an observational study.
- Because researchers in the text example first identified subjects who studied music and then collected data on their past grades, this was a retrospective study.
- Had the researchers identified subjects in advance and collected data as events unfolded, the study would have been a prospective study.
- Observational studies are valuable for discovering trends and possible relationships.
- However, it is not possible for observational studies to demonstrate a causal relationship.

Randomized, Comparative Experiments

- An experiment is a study design that allows us to prove a cause-and-effect relationship.
- An experiment:
    - *Manipulates* factor levels to create treatments.
    - *Randomly* assigns subjects to these treatment levels.
    - *Compares* the responses of the subject groups across treatment levels.
- In an experiment, the experimenter must identify at least one explanatory variable, called a factor, to manipulate and at least one response variable to measure.

Randomized, Comparative Experiments (cont.)

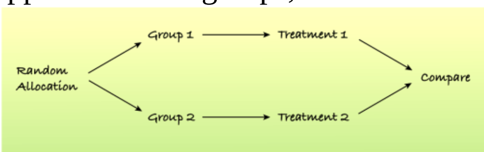
- In an experiment, the experimenter actively and deliberately manipulates the factors to control the details of the possible treatments, and assigns the subjects to those treatments *at random*.
- The experimenter then observes the response variable and *compares* responses for different groups of subjects who have been treated differently.
- In general, the individuals on whom or which we experiment are called experimental units.
    - When humans are involved, they are commonly called subjects or participants.
- The specific values that the experimenter chooses for a factor are called the levels of the factor.
- A treatment is a combination of specific levels from all the factors that an experimental unit receives.

The Four Principles of Experimental Design

1. Control:
    - We control sources of variation other than the factors we are testing by making conditions as similar as possible for all treatment groups.
2. Randomize:
    - Randomization allows us to equalize the effects of unknown or uncontrollable sources of variation.
        - It does not eliminate the effects of these sources, but it spreads them out across the treatment levels so that we can see past them.
    - Without randomization, you do not have a valid experiment and will not be able to use the powerful methods of Statistics to draw conclusions from your study.
3. Replicate:
    - Repeat the experiment, applying the treatments to a number of subjects.
        - The outcome of an experiment on a single subject is an anecdote, not data.
    - When the experimental group is not a representative sample of the population of interest, we might want to replicate an entire experiment for different groups, in different situations, etc.
4. Block:
    - Sometimes, attributes of the experimental units that we are not studying and that we can't control may nevertheless affect the outcomes of an experiment.
    - If we group similar individuals together and then randomize within each of these blocks, we can remove much of the variability due to the difference among the blocks.
    - Note: Blocking is an important compromise between randomization and control, but, unlike the first three principles, is not *required* in an experimental design.

Diagrams of Experiments

- It's often helpful to diagram the procedure of an experiment.
- The following diagram emphasizes the random allocation of subjects to treatment groups, the separate treatments applied to these groups, and the ultimate comparison of results:



Does the Difference Make a Difference?

- How large do the differences need to be to say that there is a difference in the treatments?
- Differences that are larger than we'd get just from the randomization alone are called statistically significant.
- We'll talk more about statistical significance later on. For now, the important point is that a difference is statistically significant if we don't believe that it's likely to have occurred only by chance.

Experiments and Samples

- Both experiments and sample surveys use randomization to get unbiased data.
- But they do so in different ways and for different purposes:
  - Sample surveys try to estimate population parameters, so the sample needs to be as representative of the population as possible.
  - Experiments try to assess the effects of treatments, and experimental units are not always drawn randomly from a population.

Control Treatments

- Often, we want to *compare* a situation involving a specific treatment to the status quo situation.
- A baseline ("business as usual") measurement is called a control treatment, and the experimental units to whom it is applied is called the control group.

Blinding

- When we know what treatment was assigned, it's difficult not to let that knowledge influence our assessment of the response, even when we try to be careful.
- In order to avoid the bias that might result from knowing what treatment was assigned, we use blinding.
- There are two main classes of individuals who can affect the outcome of the experiment:
  - those who could influence the results (subjects, treatment administrators, technicians)
  - those who evaluate the results (judges, treating physicians, etc.)
- When every individual in *either one* of these classes is blinded, an experiment is said to be single-blind.
- When everyone in *both* classes is blinded, the experiment is called double-blind.

Placebos

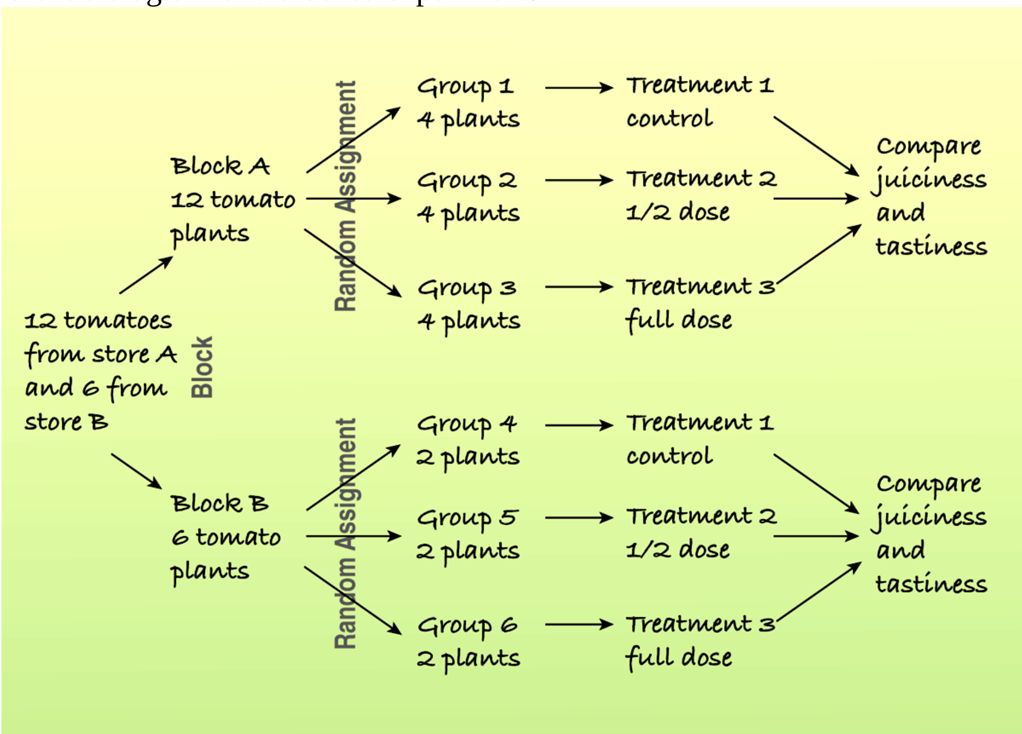
- Often simply applying *any* treatment can induce an improvement.
- To separate out the effects of the treatment of interest, we can use a control treatment that mimics the treatment itself.
- A "fake" treatment that looks just like the treatment being tested is called a placebo.
  - Placebos are the best way to blind subjects from knowing whether they are receiving the treatment or not.
- The placebo effect occurs when taking the sham treatment results in a change in the response variable.
  - This highlights both the importance of effective blinding and the importance of comparing treatments with a control.
- Placebo controls are so effective that you should use them as an essential tool for blinding whenever possible.

The Best Experiments…

- are usually:
    - randomized.
    - comparative.
    - double-blind.
    - placebo-controlled.

Blocking

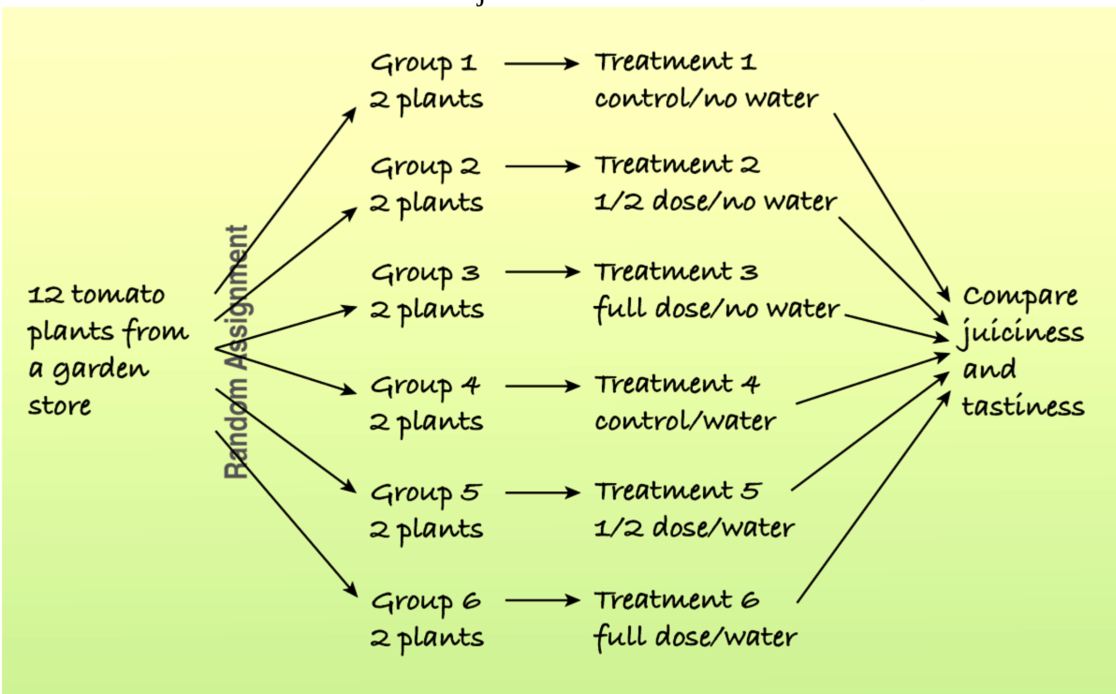
- When groups of experimental units are similar, it's often a good idea to gather them together into blocks.
- Blocking isolates the variability due to the differences between the blocks so that we can see the differences due to the treatments more clearly.
- When randomization occurs only within the blocks, we call the design a randomized block design.
- Here is a diagram of a blocked experiment:



- In a retrospective or prospective study, subjects are sometimes paired because they are similar in ways *not* under study.
    - Matching subjects in this way can reduce variability in much the same way as blocking.
- Blocking is the same idea for experiments as stratifying is for sampling.
    - Both methods group together subjects that are similar and randomize within those groups as a way to remove unwanted variation.
    - We use blocks to reduce variability so we can see the effects of the factors; we're not usually interested in studying the effects of the blocks themselves.

*Adding More Factors

- It is often important to include multiple factors in the same experiment in order to examine what happens when the factor levels are applied in different *combinations*.
- For example, the following diagram shows a study of the effects of different fertilizer/water combinations on the juiciness and tastiness of tomatoes:



Confounding

- When the levels of one factor are associated with the levels of another factor, we say that these two factors are confounded.
- When we have confounded factors, we cannot separate out the effects of one factor from the effects of the other factor.

Lurking or Confounding

- A lurking variable creates an association between two other variables that tempts us to think that one may cause the other.
  - This can happen in a regression analysis or an observational study.
  - A lurking variable is usually thought of as a prior cause of both *y* and *x* that makes it appear that *x* may be causing *y*.
- Confounding can arise in experiments when some other variables associated with a factor has an effect on the response variable.
  - Since the experimenter *assigns* treatments (at random) to subjects rather than just observing them, a confounding variable can't be thought of as causing that assignment.
- A confounding variable, then, is associated in a noncausal way with a factor and affects the response.
  - Because of the confounding, we find that we can't tell whether any effect we see was caused by our factor or by the confounding factor (or by both working together).

What Can Go Wrong?

- Don't give up just because you can't run an experiment.
  - If we can't run an experiment, often an observational study is a good choice.
- Beware of confounding.
  - Use randomization whenever possible to ensure that the factors not in you experiment are not confounded with your treatment levels.
  - Be alert to confounding that cannot be avoided, and report it along with your results.
- Bad things can happen even to good experiments.
  - Protect yourself by recording additional information.
- Don't spend your entire budget on the first run.
  - Try a small pilot experiment before running the full-scale experiment.
  - You may learn some things that will help you make the full-scale experiment better.