

Chapter 1: Data Collection and Producing Data (Continued)

4. Observational Study and Experiment:

- **A Observational Study:** Measures the characteristics of a population by studying individuals in a sample, but does not attempt to manipulate or influence the variables of interest.
 - **Example 1:** Who gets good grades? And, more importantly, why? Is there something schools and parents could do to help weaker students improve their grades? Some people think they have an answer: music! No, not your portable MP3 player, but an instrument. In a 1981 study conducted at Mission Viejo High School, in California, researchers compared the scholastic performance of music students with that of non-music students. Guess what? The music students had a much higher overall grade point average than the non-music students, 3.59 to 2.91. Not only that, a whopping 16% of the music students had all A's compared with only 5% of the non-music students.

As a result of this study and others, many parent groups and educators pressed for expanded music programs in the nation's schools. They argued that the work ethic, discipline, and feeling of accomplishment fostered by learning to play an instrument also enhance a person's ability to succeed in school. They thought that involving more students in music would raise academic performance. What do you think? Does this study provide solid evidence? Or are there other possible explanations for the difference in grades? Is there any way to really prove such a conjecture?

This research tried to show an association between music education and grades. But it didn't assign students to get music education. Instead, it simply observed students "in the wild," recording the choices they made and the outcome. Such studies are called observational studies. In observational studies, researchers don't assign choices; they simply observe them.

What's wrong with concluding that music education causes good grades? One high school during one academic year may not be representative of the whole United States. That's true, but the real problem is that the claim that music study caused higher grades depends on there being no other differences between the groups that could account for the differences in grades.

We can think of lots of other reasons why the groups might perform differently. Students who study music may have better work habits to start with, and this makes them successful in both music and course work. Music students may have more parental support and that support enhanced their academic performance, too. Maybe they came from wealthier homes

and had other advantages. Or it could be that smarter kids just like to play musical instruments.

Lurking variables: the characteristics or variables that may affect the response, but that are not included in the study, are referred to as lurking variables.

This study can't be used to prove that music lessons *cause* grades to improve. How can we do better? Is it ever possible to prove a cause-and-effect relationship? Well, yes it is, but we would have to take a different approach.

Only a well designed experiment can prove a cause-and-effect relationship.

We could take a group of third graders, randomly assign half to take music lessons, and forbid the other half to do so. Then we could compare their grades several years later. This kind of study design is called an experiment. An experiment requires a random assignment of subjects to treatments.

Only an experiment can justify a claim like "music lessons cause higher grades." Questions like "Does taking vitamin C reduce the chance of getting a cold?" and "Does working with computers improve performance in Statistics class?" and "Is this drug a safe and effective treatment for that disease?" require a designed experiment to establish cause and effect.

- **Example 2:** Does smoking cause lung cancer? *(The following is adapted from Basic Practice of Statistics, 3rd edition, by D. Moore.)*

Doctors had long observed that most lung cancer patients were smokers. Comparison of smokers and "similar" nonsmokers showed a very strong association between smoking and death from lung cancer. Could the association be explained by lurking variables? Might there be, for example, a genetic factor that predisposes people both to nicotine addiction and to lung cancer? Smoking and lung cancer would then be positively associated even if smoking had no direct effect on the lungs. How were these objects overcome?

What are the criteria for establishing causation when we can not do an experiment?

- * The association between smoking and lung cancer is very strong.
- * The association is consistent. Many studies of different kinds of people in many countries link smoking to lung cancer. That reduces the chance that a lurking variable specific to one group or one study explains the association.
- * Higher doses are associated with stronger responses. People who smoke more cigarettes per day or who smoke over a longer period get lung cancer more often. People who stop smoking reduce their risk.
- * The alleged cause precedes the effect in time. Lung cancer develops

after years of smoking. The number of men dying of lung cancer rose as smoking became more common, with a lag of about 30 years. Lung cancer kills more men than any other form of cancer. Lung cancer was rare among women until women began to smoke. Lung cancer in women rose along with smoking, again with a lag of about 30 years, and has now passed breast cancer as the leading cause of cancer death women.

- * The alleged cause is plausible. Experiments with animals show that tars from cigarette smoke do cause cancer.
- A **designed experiment** applies a treatment to individuals (referred to as experimental units or subjects) and attempts to isolate the effects of treatment on a response variable. Factors are the explanatory variables that have effects on the response in the study. A treatment is any combination of the values of each factor.
 - **Three basic principles of statistical design of experiment:**
 1. Control the effects of lurking variables on the response, most simply by comparing several treatments.
 2. Randomization - use impersonal chance to assign subjects to treatments.
 3. Replication - use enough subjects in each group to reduce variation in the results.
 - A **completely randomized design** is one in which each experimental units is randomly assigned to a treatment.

Example: A farmer wishes to determine the optimal level of a new fertilizer on his soybean crop. Design an experiment that will assist him.

- * Step 1: Identifying the response: yield of soybean
- * Step 2: Identifying the factor and its levels which will affect the yield: fertilize

Treatments: (Levels of fertilizer)

- Treatment A: 20 soybean plants receive no fertilizer;
- Treatment B: 20 soybean receive 2 teaspoons of fertilizer per gallon of water every 2 weeks;
- Treatment C: 20 soybean receive 4 teaspoons of fertilizer per gallon of water every 2 weeks;

other factors: precipitation, sunlight, type of soil, etc .(Assume they are the same to the soybean.)

- * Step 3: Make a diagram of the design

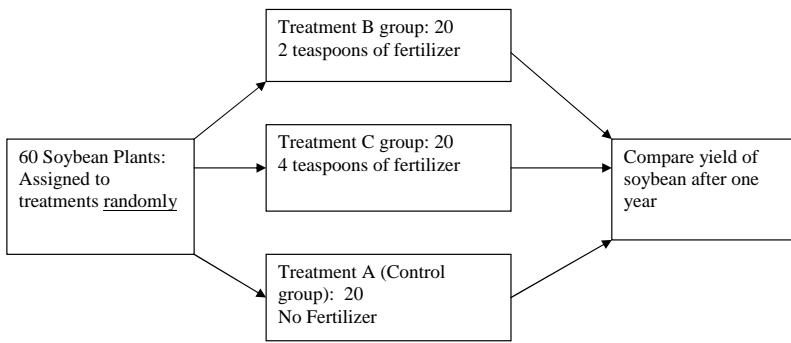


Figure 1: Randomized Comparison Experiment of Soybean Fertilizer

- **Control** is one of the basic principles of statistical design of experiment. **Placebo effect** comes from a dummy treatment. An experiment that uses both comparison and randomization is called a **randomized comparison experiment**.

* **Example:** Gastric freezing.

Gastric freezing is a clever treatment for ulcers in the upper intestine. The patient swallows a deflated balloon with tubes attached, then a refrigerated liquid is pumped through the balloon for an hour. The idea is that cooling the stomach will reduce its production of acid and so relieve ulcers. An experiment reported in the *Journal of the American Medical Association* showed that gastric freezing did reduce acid production and relieve ulcer pain. The treatment was safe and easy and was widely used for several years. The diagram of design of the experiment was

Gastric freezing → Observe pain relief

The gastric freezing experiment was poorly designed. The patients' response may have been due to the placebo effect. A placebo is a dummy treatment. Many patients respond favorably to any treatment, even a placebo, presumably because of trust on the doctor and expectations of a cure. This response to dummy treatment is the placebo effect.

A later experiment divided ulcer patients into two groups. One group was treated by gastric freezing as before. The other group received a placebo treatment in which the liquid in the balloon was at body temperature rather than freezing. The results: 34% of the 82 patients in the treatment group improved, but so did 38% of the 78 patients in the placebo group. This and other properly designed experiments showed that gastric freezing was no better than a placebo, and its use was abandoned.

The first gastric freezing experiment gave misleading results because the effects of the explanatory variable (gastric freezing) were confounded

Randomized Comparison Experiment: Gastric Freezing

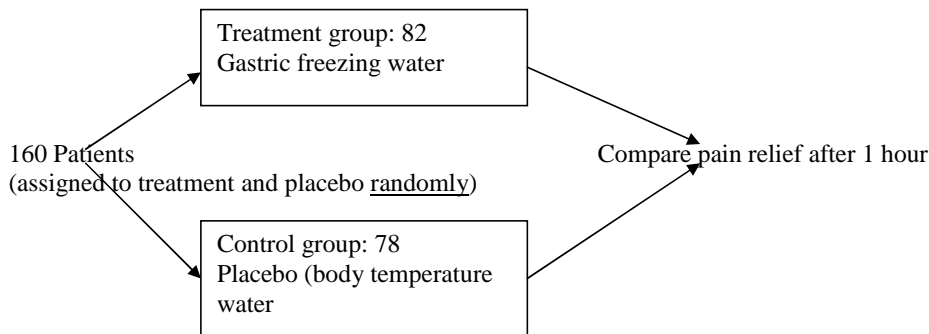


Figure 2: Randomized Comparison Experiment of Gastric Freezing

with (mixed up with) the placebo effect. We can defeat confounding by comparing two groups of patients, as in the second gastric freezing experiment. The placebo effect and other lurking variables now operate on both groups. The only difference between the groups is the actual effect of gastric freezing. The group of patients who received a sham treatment is called a control group, because it enables us to control the effects of outside variables on the outcome. Control is one of three basic principle of statistical design of experiments. Comparison of several treatments in the same environment is the simplest form of control.

Without control, experimental results in medicine and the behavioral sciences can be dominated by such influences as the details of the experimental arrangement, the selection of subject, and the placebo effect. The result is often biased. The design of a study is biased if it systematically favors certain outcomes.

An uncontrolled study of a new medical therapy is biased in favor of finding the treatment effective because of the placebo effect. It should not surprise you to learn that uncontrolled studies in medicine give new therapies a much higher success rate than proper comparative experiments. Well-designed experiments, like the second gastric freezing study and soybean fertilizer, usually compare several treatments.

– **Double-Blind Experiments:**

- * Placebo works. That bare fact means that medical studies must take special care to show that a new treatment is not just a placebo. Part of equal treatment for all is to be sure that the placebo effect operates on all subjects.

- * The Powerful Placebo:

Example 1: Want to help balding men keep their hair? Give them a placebo- one study found that 42% of balding men maintained or increased the amount of hair on their heads when they took a placebo.

Example 2: Another study told 13 people who were very sensitive to poison ivy that the stuff being rubbed on one arm was poison ivy. In fact, it was a placebo, but all 13 broke out in a rash. The stuff rubbed on the other arm really was poison ivy, but the persons were told it was harmless- and only 2 of 13 developed a rash.

When the ailment is vague and psychological, like depression, some experts think that about three-quarters of the effect of the most widely used drugs is just the placebo effect. Others disagree. The strength of the placebo effect in medical treatment is hard to pin down because it depends on the exact environment, much as the behavior of the fickle mice do. How enthusiastic the doctor is seems to matter a lot. But "placebos work" is a good place to start when you think about planning medical experiments.

The strength of the placebo effect is a strong argument for randomized comparative experiments.

Example 3: In the baldness study, 42% of the placebo group kept or increased their hair, but 86% of the men getting a new drug to fight baldness did so. The drug beats the placebo, so it has something besides the placebo going for it. Of course, the placebo effect is still part of the reason this and other treatments work.

Because the placebo effect is so strong, it would be foolish to tell subjects in a medical experiment whether they are receiving a new drug or a placebo. Knowing that they are getting "just a placebo" might weaken the placebo effect and bias the experiment in favor of the other treatments. It is also foolish to tell doctors and other medical personnel what treatment each subject received. If they know that a subject is getting "just a placebo", they may expect less than if they know the subject is receiving a "promising" experimental drug. Doctors' expectations change how they interact with patients and even the way they diagnose a patient's condition. Whenever possible, experiments with human subjects should be double-blind.

In a **double-blind experiment**, neither the subjects nor the people who work with them know which treatment each subject is receiving.

Until the study ends and the results are in, only the study's statistician knows for sure. Reports in medical journals regularly begin with words like these, from a study of a flu vaccine given as a nose spray: "This study was a randomized, double-blind, placebo-controlled trial. Participants were enrolled from 13 sites across the continental United States between mid-September and mid-November 1997." Doctors are supposed to know what this means. Now you also know.

Homework:

When it was found that hydroxyurea reduced the symptoms of sickle cell anemia, the National Institutes of Health released a medical bulletin. The bulletin said: "These findings are the results of data analyzed from the Multi-center Study of Hydroxyurea in Sickle Cell Anemia (MSH), which was a double-blind, placebo-controlled trial in placebo capsule." Explain to someone who knows no statistics what the terms "placebo-controlled" and "double-blind" mean here.