In contrast to observational studies, experiments don't just observe individuals or ask them questions. They actively impose some treatment in order to measure the response.

An **observational study** observes individuals and measures variables of interest but does not attempt to influence the responses. The purpose is to describe some group or situation.

An **experiment** deliberately imposes some treatment on individuals to measure their responses. The purpose is to study whether the treatment causes a change in the response.

When our goal is to understand **cause and effect**, experiments are the *only* source of fully convincing data.

The distinction between observational study and experiment is one of the most important in statistics.

An experiment is a statistical study in which we actually do something (a **treatment**) to people, animals, or objects (the **experimental units**) to observe the **response.** Here is the basic vocabulary of experiments.

The **experimental units** are the smallest collection of individuals to which treatments are applied. When the units are human beings, they often are called **subjects.**

The explanatory variables in an experiment are often called **factors.**

A specific condition applied to the individuals in an experiment is called a **treatment.** If an experiment has several explanatory variables, a treatment is a combination of specific values of these variables.

Ex. Cell phones and brain cancer. A study of cell phones and the risk of brain cancer looked at a group of 469 people who have brain cancer. The investigators matched each cancer patient with a person of the same sex, age, and race who did not have brain cancer, then asked about use of cell phones. Result: “Our data suggest that use of handheld cellular telephones is not associated with risk of brain cancer.”

What are the explanatory and response variables?
Is this an observational study or an experiment? Why?

This is an observational study.
No treatment was assigned to the subjects; we merely observed cell phone usage (and presence/absence of cancer).
Explanatory variable: cell phone usage;
Response variable: whether or not a subject has brain cancer.

Ex. Do abandoned children placed in foster homes do better than similar children placed in an institution?

The Bucharest Early Intervention Project found that the answer is a clear "Yes." The subjects were 136 young children abandoned at birth and living in orphanages in Bucharest, Romania. Half of the children, chosen at random, were placed in foster homes (paid for by the study). The other half remained in the orphanages.

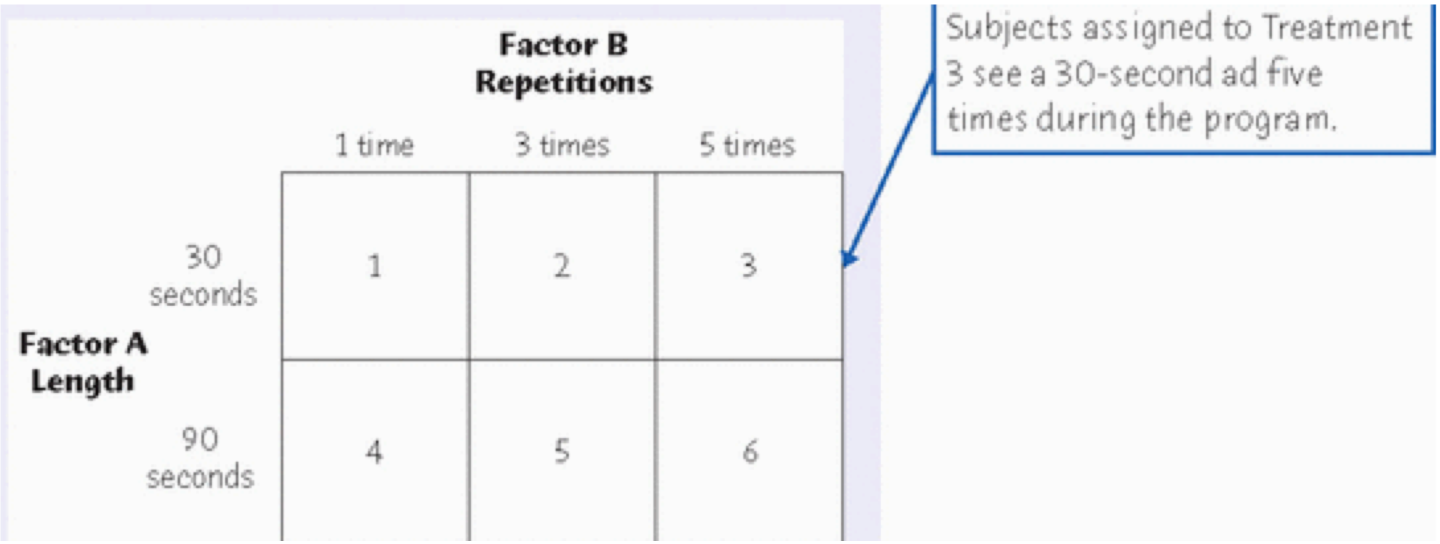
What are the explanatory and response variables?
Is this an observational study or an experiment? Why?

The experiment compared these two treatments. There is a single factor, type of care (foster versus institutional care). The response variables included measures of mental and physical development.

Ex. What are the effects of watching an ad on TV?

The answer may depend both on the length of the ad and on how often it is repeated. An experiment investigated this question using undergraduate students as subjects. All subjects viewed a 40-minute television program that included ads for a digital camera. Some subjects saw a 30-second commercial; others, a 90-second version. The same commercial was shown either 1, 3, or 5 times during the program.

This experiment has **2 factors**: length of the commercial, with 2 values, and repetitions, with 3 values. The **6 combinations** of 1 value of each factor form **6 treatments**. Figure shows the layout of the treatments. After viewing, all the subjects answered questions about their recall of the ad, their attitude toward the camera, and their intention to purchase it. These are the **response** variables.



Ex. Growing in the shade. Ability to grow in shade may help pines in the dry forests of Arizona resist drought. How well do these pines grow in shade? Plant pine seedlings in a greenhouse in either full light, light reduced to 25% of normal by shade cloth, or light reduced to 5% of normal. At the end of the study, dry the young trees and weigh them.

Individuals: pine seedlings.
Factor: amount of light.
Treatments: full light, 25% light, or 5% light.
Response variable: dry weight at the end of the study.

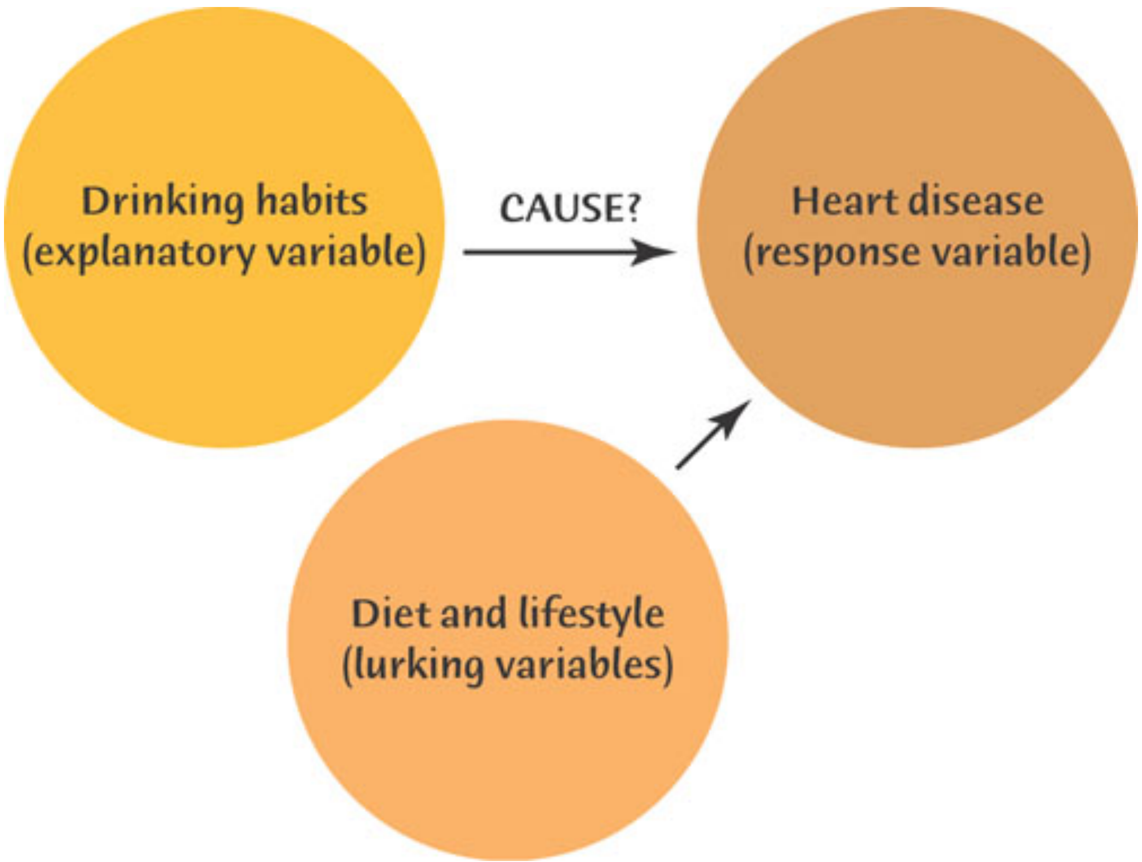
Observational studies of the effect of one variable on another often fail because of **confounding** between the explanatory variable and one or more **lurking variables.**

A **lurking variable** is a variable that is not among the explanatory or response variables in a study but that may influence the response variable.

Confounding occurs when two variables are associated in such a way that their effects on a response variable cannot be distinguished from each other.

Well-designed experiments take steps to avoid confounding.

Variables are confounded when their effects on a response can't be distinguished from each other. Observational studies and uncontrolled experiments often fail to show that changes in an explanatory variable actually cause changes in a response variable because the explanatory variable is confounded with lurking variables.



Ex. Is the online course more effective effective than the classroom version?

This experiment has a very simple design. A group of subjects (the students) were exposed to a treatment (the online course), and the outcome (GMAT scores) was observed.
Here is the design:

Subjects → Online course → GMAT scores
The students in the online review course were different from the students who in past years took the classroom course—they were older and more likely to be employed. We can't compare the performance of these mature students with that of the undergraduates who previously dominated the course. The online course might even be less effective than the classroom version.

The effect of online versus in-class instruction is **confounded** with the effect of lurking variables.

Uncontrolled experiments often yield worthless results because of confounding with lurking variables

Randomized comparative experiments are designed to give good evidence that differences in the treatments actually cause the differences we see in the response.

Principles of Experimental Design

1. **Control** for lurking variables that might affect the response, most simply by comparing two or more treatments.
2. **Randomize:** Use chance to assign experimental units to treatments.
3. **Replication:** Use enough experimental units in each group to reduce chance variation in the results.

The logic of a randomized comparative experiment depends on our ability to treat all the subjects the same in every way except for the actual treatments being compared.

A **placebo** is a dummy treatment. Experiments in medicine and psychology often give a placebo to a control group because just being in an experiment can affect responses.

In a **double-blind experiment,** neither the subjects nor those who interact with them and measure the response variable know which treatment a subject received.

Ex. Will daily doses of vitamin E reduce the risk of heart disease?

Divide the subjects at random into two groups. All the subjects receive the same medical attention during the several years of the experiment. All of them take a pill every day: vitamin E in the treatment group and a dummy pill (called a placebo) in the control group. Many patients respond favorably to any treatment, even a placebo, perhaps because they trust the doctor.

The response to a dummy treatment is called **the placebo effect**.

If the control group did not take any pills, the effect of vitamin E in the treatment group would be confounded with the placebo effect, the effect of simply taking pills.

In addition, the experiment **is double–blind**. The subjects don't know whether they are taking vitamin E or a placebo. Neither do the medical personnel who work with them. The double–blind method avoids unconscious bias by, for example, a doctor who is convinced that a vitamin must be better than a placebo.

Good experiments require attention to detail as well as good statistical design. Many behavioral and medical experiments are **double–blind.** Some give a **placebo** to a control group.

Even well-designed experiments often face another problem**: lack of realism**. It can limit our ability to apply the conclusions of an experiment to the settings of greatest interest.

Ex. Lack of realism

The study of television advertising in previous example showed a 40-minute video to students who knew an experiment was going on. We can't be sure that the results apply to everyday television viewers. Many behavioral science experiments use as subjects students or other volunteers who know they are subjects in an experiment. That's not a realistic setting.

Matched Pairs Designs

A common type of randomized design for comparing two treatments is a matched pairs design. The idea is to create blocks by matching pairs of similar experimental units.

A **matched pairs design** compares two treatments. Choose pairs of subjects that are as closely matched as possible. Use chance to decide which subject in a pair gets the first treatment. The other subject in that pair gets the other treatment.

Sometimes, a "pair" in a matched pairs design consists of a single unit that receives both treatments. Since the order of the treatments can influence the response, chance is used to determine which treatment is applied first for each unit.

Ex. Does talking on a hands-free cell phone distract drivers? Undergraduate students "drove" in a high-fidelity driving simulator equipped with a hands-free cell phone. The car ahead brakes: how quickly does the subject react?

Let's compare two designs for this experiment. There are 40 student subjects available.

In a **completely randomized design**, all 40 subjects are assigned at random, 20 to simply drive and the other 20 to talk on the cell phone while driving. In the matched pairs design that was actually used, all subjects drive both with and without using the cell phone. The two drives are on separate days to reduce carryover effects. The order of the two treatments is assigned at random: 20 subjects are chosen to drive first with the phone, and the remaining 20 drive first without the phone.

Some subjects naturally react faster than others. **The completely randomized design** relies on chance to distribute the faster subjects roughly evenly between the two groups. **The matched pairs design** compares each subject's reaction time with and without the cell phone. This makes it easier to see the effects of using the phone.

Ex. How long did I work? A psychologist wants to know if the difficulty of a task influences our estimate of how long we spend working at it. She designs two sets of mazes that subjects can work through on a computer. One set has easy mazes and the other has hard mazes. Subjects work until told to stop. They are then asked to estimate how long they worked. The psychologist has 30 students available to serve as subjects.

(a) Describe the design of a completely randomized experiment to learn the effect of difficulty on estimated time.
(b) Describe the design of a matched pairs experiment using the same 30 subjects.

(a) **Completely randomized design**: Randomly assign 15 students to Group 1 (easy mazes) and the other 15 to Group 2 (hard mazes). Compare the time estimates of Group 1 with those of Group 2.
(b) **Matched pairs design**: Each student does the activity twice, once with the easy mazes and once with the hard mazes. Randomly decide (for each student) which set of mazes will be used first. Compare each student's "easy" and "hard" time estimate (for example, by looking at each "hard" minus "easy" difference).