1. DEs and OSs

Chapters 1 and 2 contrast designed experiments (DEs) with observational studies (OSs).

The usual purpose of this kind of research is usually to draw conclusions about causation. For example,

- Does the death penalty deter crime? More police?
- Does education make you richer?
- Does airport security decrease terrorism? (Is terrorism increasing globally year over year?)

A double-blind, randomized, controlled, cross-over experiment usually gives more accurate conclusions than an observational study.

Designed Experiments

The ideal for a statistical study is the double-blind, randomized, controlled cross-over experiment.

A study is controlled if one group receives the treatment and another group does not. (In medicine, that group usually gets either a placebo, or standard medical care, or both. Sometimes pairing, e.g. for socio-economic class.)

A study is double-blind if neither the subjects nor the scientists know who is assigned to which group until after the data are collected. This

- prevents subjects in different groups from behaving in different ways;
- prevents scientists from introducing any unconcious bias into the data collection process.

A study is randomized if the control group and the treatment group are chosen at random.

Without randomization, the groups may differ in a systematic way. For example, surgeons used to assign only the healthiest patients to receive an experimental new surgical treatment, since those patients could best withstand the invasive procedure. But the outcomes for those patients is not a reliable forecast for how normal patients would respond. cf. Sex bias, p. 17.

Historical controls do not give a randomized experiment. Why?

Cross-over: The treatments (placebo included) are given in random order with repeats.

Statisticians are the watchdogs....

- Most accidents occur within 2 miles of home. So, it's much safer to drive transcontinentally than to Whole foods.
- Salaries of Presbyterian ministers in NH have been increasing. So has the amount of Cuban rum illegally imported into Florida.
- Crime is reported more and more. It must be going up. Buy a gun.
- Fertility rates peaked in 1957. The birth rate peaked in 1963!?
- Callers to a talk show were 7 to 1 against divorce. Time to make it illegal.

We ask the sensitive questions....Sex, race, crime, money, religion, mental health...You don't know what the reality is until you collect and analyze data correctly. Midtown Man./Nova Scotia/Uniform prevalence.

Designing an Experiment

Advocates of natural medicine have a herbal tea they think will help restore reduced function in elderly. You have access to some senior care facilities. Say you have a budget to test up to 30 subjects.

What do you do?

Later in the course we will learn how to decide what level of success is significant. 18 successes out of 30 is not surprising, but 28 out of 30 (probably) provides strong evidence.

Class discussion questions: Explain whether this experiment is:

- double-blind
- randomized
- controlled
- cross-over

Suppose the treatment was acupuncture rather than herbal tea.

If you could not find 30 volunteers, could you still perform this experiment with a smaller sample size?

Observational Studies

In an observational study, the researcher does not get to determine who recieves the treatment.

For example, people who smoke get lung cancer at a higher rate than those who do not smoke. Does smoking cause lung cancer?

The tobacco lobby used to say no, arguing that

- there might be a gene that predisposes people to both enjoy smoking and get cancer;
- people who like to smoke may tend to follow unhealthy lifestyles (e.g., less exercise), and that may be the real cause of lung cancer;
- no randomized, controlled, double-blind experiment (on humans) had been performed to show causation.

It would be unethical to do a randomized controlled experiment – one would have to assign 14 year-olds at random to smoke heavily for the rest of their lives. And it would be hard to make this double-blind – people know if they smoke.

But animal studies strongly indicate that smoking causes lung cancer in mammals and birds.

The other two arguments need refutation. The differences between lung cancer rates in the smokers and non-smokers may be due to smoking, or they may be due to a confounding factor.

In this case, two possible confounding factors are genes and lifestyle.

A confounding factor is associated with both

- outcome
- group membership.

For example, one might argue that lung cancer is caused by matches, not tobacco. Or one might argue that cholesterol does not cause heart disease, but rather is a result of poor circulation or breakdown of heart muscle tissue.

One way to try to handle confounding is to make subgroup comparisons that control for possible confounding effects. For example, one could compare the lung cancer rates for smokers who use matches against smokers who use lighters.

Do seatbelts save lives?

Seatbelt studies are usually observational (why?). One compares the fatality rates in accidents in which seatbelts were worn to the fatality rate in accidents without seatbelts.

But one has to worry about confounding factors. For example,

- People who don't wear seatbelts may tend to drive faster.
- People who don't wear seatbelts may tend to drive cars that are not designed with safety in mind.

So people try to control for this by comparing the fatality rates among seatbelt wearers and non-wearers in similar cars, or cars that are thought to have been traveling at the same speed. But this is awkward and invites criticism.

Educational testing (Mismeasure of Man, Gould). One picture shows a man reading. The other picture shows a man chopping wood. Which man is working?

Educational testing questions often have a passage to read and then ask some questions. How fair is this? Item parameters, subject parameters...

In order to control for a confounding factor, one has to guess what it is. That can be hard; you can't prove you have thought of everything.

In contrast, with a randomized design, the random assignment of people to the treatment and control groups ensures that there is almost no chance of a systematic difference between the groups. You are unlikely to get most of the safe drivers in one group and the reckless in the other, or most of the people with good genes for lung cancer in one group and all those with bad genes in the other.

Some people say that prayer/positive thinking/other psychological activities promote health. How would you investigate the STATISTICAL validity of the claim? How would you assess improvement in a disease as a consequence of frame of mind?