Statistics: Concepts and Controversies

Chapter 6 Experiments in the Real World

Lecture Slides

Case Study: Experiments in the Real World (1 of 4)

Is caffeine dependence real?

Researchers at Johns Hopkins Univ. School of Medicine studied whether some people develop a serious addiction called caffeine dependence syndrome.

Photononstop/SuperStock

Case Study: Experiments in the Real World (2 of 4)

Eleven caffeine dependent volunteers took either their daily amount of caffeine or a fake (nonactive substance) for two days.

Randomization was used to determine whether a subject took their daily caffeine or the fake substance.

Photononstop/SuperStock

Case Study: Experiments in the Real World (3 of 4)

At least a week later, subjects received the opposite treatment for a two-day period. (If a subject took caffeine during the first two-day period, they took the fake substance during the second two-day period and vice versa.)

Subjects' diets were restricted during the study periods. They were not allowed caffeine and also were not allowed products with artificial sweeteners (this hopefully diverted their attention from caffeine).

Case Study: Experiments in the Real World (4 of 4)

Several assessments were administered at the end of each two-day period.

- Questionnaires assessed mood, depression, and the presence of certain physical symptoms.
- Subjects were given a tapping task where they were asked to press a button 200 times as fast as they could.
- Researchers, who did not know whether subjects took caffeine or the fake substance, interviewed subjects to look for signs of functional impairment.

By the end of this chapter, you be able to determine the strengths and weaknesses of a study such as this.

Equal Treatments for All

Probability samples are a good start, but sampling in practice has difficulties that just using a random sample won't solve.

Using a randomized comparative experiment is also a good idea, but it doesn't solve all the difficulties of experimenting.

The logic of a randomized comparative experiment assumes that all the subjects are treated alike except for the treatments that the experiment is designed to compare. Any other unequal treatment can cause bias.

Example: Rats and Rabbits (1 of 2)

Does a new breakfast cereal provide good nutrition? To find out, compare the weight gains of young rats fed the new product and rats fed a standard diet.

Rats are randomly assigned to diets and are housed in large racks of cages.

It turns out that rats in upper cages grow a bit faster than rats in bottom cages.

If the experimenters put rats fed the new product at the top and those fed the standard diet below, the experiment is biased in favor of the new product.

Solution: assign the rats to cages at random.

Example: Rats and Rabbits (2 of 2)

Another study looked at the effects of human affection on the cholesterol level of rabbits.

All the rabbit subjects ate the same diet.

Some (chosen at random) were regularly removed from their cages to have their furry heads scratched by friendly people.

The rabbits who received affection had lower cholesterol. So affection for some but not other rabbits could bias an experiment in which the rabbits' cholesterol levels is a response variable.

Double-blind Experiments (1 of 3)

Placebos “work.”

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.

Example: The Powerful Placebo (1 of 2)

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.

Another study told 13 people who were very sensitive to poison ivy that the stuff being rubbed on one arm was poison ivy. 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 subjects were told it was harmless—and only 2 of the 13 developed a rash.

Example: The Powerful Placebo (2 of 2)

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 treatments is hard to pin down because it depends on the exact environment. 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.

Double-blind Experiments (2 of 3)

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 is receiving. 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.

Double-blind Experiments (3 of 3)

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.

Refusals, Nonadherers, and Dropouts

Sample surveys suffer from nonresponse due to failure to contact some people selected for the sample and the refusal of others to participate.

Experiments with human subjects suffer from similar problems. Nonadherers also impact the interpretation of results by not following the specified protocol or by dropping out of the experiment.

Example: Minorities in Clinical Trials

The law now requires representation of women and minorities, and data show that most clinical trials now have fair representation. Refusals remain a problem. Minorities, especially blacks, are more likely to refuse to participate.

“A major impediment for lack of participation is a lack of trust in the medical establishment.”

Some remedies are complete and clear information, insurance coverage, participation of black researchers, and cooperation with doctors and health organizations in black communities.

Example: Dropouts in a Medical Study (1 of 2)

Orlistat blocks absorption of fat from the foods we eat. The drug was compared with a placebo in a double-blind randomized trial with 1187 obese subjects.

They were given a placebo for four weeks, and the subjects who wouldn't take a pill regularly were dropped. This addressed the problem of nonadherers.

There were 892 subjects left. These subjects were randomly assigned to orlistat or a placebo, along with a weight-loss diet.

Example: Dropouts in a Medical Study (2 of 2)

After a year devoted to losing weight, 576 subjects were still participating.

On average, the orlistat group lost 3.15 kilograms (about 7 pounds) more than the placebo group.

At the end of the second year, 403 subjects were left. That's only 45% of the 892 who were randomized. Orlistat again beat the placebo, reducing the weight regained by an average of 2.25 kilograms (about 5 pounds).

Can We Generalize?

A well-designed experiment tells us that changes in the explanatory variable cause changes in the response variable.

Can we generalize our conclusions from our little group of subjects to a wider population? The first step is to be sure that our findings are statistically significant, that they are too strong to often occur just by chance.

The serious threat is that the treatments, the subjects, or the environment of our experiment may not be realistic.

Example: Studying Frustration (1 of 3)

A psychologist wants to study the effects of failure and frustration on the relationships among members of a work team.

She forms a team of students, brings them to the psychology laboratory, and has them play a game that requires teamwork.

The game is rigged so that they lose regularly. The psychologist observes the students through a one-way window and notes the changes in their behavior during an evening of game playing.

Example: Studying Frustration (2 of 3)

Playing a game in a laboratory for small stakes, knowing that the session will soon be over, is a long way from working for months developing a new product that never works right and is finally abandoned by your company.

Does the behavior of the students in the lab tell us much about the behavior of the team whose product failed?

Example: Studying Frustration (3 of 3)

The subjects (students who know they are subjects in an experiment), the treatment (a rigged game), and the environment (the psychology lab) are all unrealistic if the psychologist's goal is to reach conclusions about the effects of frustration on teamwork in the workplace. Psychologists do their best to devise realistic experiments for studying human behavior, but lack of realism limits the ability to generalize beyond the environment and subjects in their study, and hence the usefulness of some experiments in this area.

Experimental Design in the Real World

In a **completely randomized experimental design**, all the experimental subjects are allocated at random among all the treatments.

A completely randomized design can have any number of explanatory variables.

Example: Can Low-fat Food Labels Lead to Obesity? (1 of 4)

What are the effects of low-fat food labels on food consumption? Do people eat more of a snack food when the food is labeled as low-fat? The answer may depend both on whether the snack food is labeled low-fat and whether the label includes serving-size information.

An experiment investigated this question using university staff, graduate students, and undergraduate students at a large university as subjects.

Example: Can Low-fat Food Labels Lead to Obesity? (2 of 4)

Over 10 late-afternoon sessions, all subjects viewed episodes of a 60-minute, made-for-television program in a theater on campus and were asked to rate the episodes.

They were also told that because it was late in the afternoon, they would be given a cold 24-ounce bottle of water and a bag of granola from a respected campus restaurant called The Spice Box. They were told to enjoy as much or as little of it as they wanted.

Example: Can Low-fat Food Labels Lead to Obesity? (3 of 4)

- Each participant received 640 calories (160 grams) of granola in a clear plastic bag that was labeled with an attractive 3.25-×-4-inch color label.
- Depending on the condition randomly assigned to the subjects, the bags were labeled either “Regular Rocky Mountain Granola” or “Low-Fat Rocky Mountain Granola.” Below this, the label indicated “Contains 1 Serving” or “Contains 2 Servings,” or it provided no serving-size information.
- As participants left the theater, they were asked how many serving sizes they believed their package contained. Out of sight of the participants, the researchers also weighed each granola bag.

Example: Can Low-fat Food Labels Lead to Obesity? (4 of 4)

Participants' statements about serving size and the actual weights of the granola bags are the response variables.

This experiment has two explanatory variables: fat content, with 2 levels, and serving size, with 3 levels.

The 6 combinations of 1 level of each variable form 6 treatments.

Matched Pairs and Block Designs (1 of 6)

Completely randomized designs are the simplest statistical designs for experiments.

Completely randomized designs are often inferior to more elaborate statistical designs.

Matching the subjects in various ways can produce more precise results than simple randomization.

Matched Pairs and Block Designs (2 of 6)

A common design that combines matching with randomization is the matched pairs design.

A **matched pairs design** compares just two treatments. Choose pairs of subjects that are as closely matched as possible.

Assign one of the treatments to each subject in a pair by tossing a coin or reading odd and even digits from Table A.

Example: Testing Insect Repellents (1 of 2)

Consumers Reports describes a method for comparing the effectiveness of two insect repellents. The active ingredient in one is 15% Deet. The active ingredient in the other is oil of lemon eucalyptus.

For each volunteer, the left arm is sprayed with one of the repellents and the right arm with the other.

This is a matched pairs design in which each subject compares two insect repellents.

Example: Testing Insect Repellents (2 of 2)

Deciding which arm receives which repellent is determined randomly.

Beginning 30 minutes after applying the repellents, once every hour volunteers put each arm in separate 8 cubic foot cages containing 200 disease-free female mosquitoes in need of a blood meal to lay their eggs.

Volunteers leave their arms in the cages for 5 minutes. The repellent is considered to have failed if a volunteer is bitten two or more times in a 5-minute session. The response is the number of one-hour sessions until a repellent fails.

Matched Pairs and Block Designs (3 of 6)

Matched pairs designs use the principles of comparison of treatments and randomization.

However, the randomization is not complete—we do not randomly assign all the subjects at once to the two treatments. Instead, we randomize only within each matched pair.

This allows matching to reduce the effect of variation among the subjects. Matched pairs are an example of block designs.

Matched Pairs and Block Designs (4 of 6)

A **block** is a group of experimental subjects that are known before the experiment to be similar in some way that is expected to affect the response to the treatments.

In a **block design**, the random assignment of subjects to treatments is carried out separately within each block.

Matched Pairs and Block Designs (5 of 6)

A block design combines the idea of creating equivalent treatment groups by matching with the principle of forming treatment groups at random. Blocks are another form of control.

They control the effects of some outside variables by bringing those variables into the experiment to form the blocks.

Example: Men, Women, and Advertising (1 of 2)

Women and men respond differently to advertising. An experiment to compare the effectiveness of three television commercials for the same product will want to look separately at the reactions of men and women, as well as assess the overall response to the ads.

A completely randomized design considers all subjects, both men and women, as a single pool. The randomization assigns subjects to three treatment groups without regard to their sex. This ignores the differences between men and women.

Example: Men, Women, and Advertising (2 of 2)

A better design considers women and men separately. Randomly assign the women to three groups, one to view each commercial. Then separately assign the men at random to three groups.

Matched Pairs and Block Designs (6 of 6)

Blocks allow us to draw separate conclusions about each block.

Blocking also allows more precise overall conclusions, because the systematic differences between men and women can be removed when we study the overall effects of the three commercials.

A wise experimenter will form blocks based on the most important unavoidable sources of variability among the experimental subjects. Randomization will then average out the effects of the remaining variation and allow an unbiased comparison of the treatments.

Statistics in Summary (1 of 3)

Because the **placebo effect** is strong, **clinical trials** and other experiments with human subjects should be **double-blind** whenever this is possible.

The double-blind method helps achieve a basic requirement of comparative experiments: **equal treatment for all subjects** except for the actual treatments the experiment is comparing.

Statistics in Summary (2 of 3)

The most common weakness in experiments is that we can't **generalize** the conclusions widely. Some experiments apply unrealistic treatments, some use subjects from a special group such as college students, and all are performed at some specific place and time. We want to see similar experiments at other places and times to confirm important findings.

Statistics in Summary (3 of 3)

Many experiments use designs that are more complex than the basic **completely randomized design**, which divides all the subjects among all the treatments in one randomization. **Matched pairs** designs compare two treatments by giving one to each of a pair of similar subjects or by giving both to the same subject in random order. **Block** designs form blocks of similar subjects and assign treatments at random separately in each block.

The big ideas of **randomization, control,** and **adequate numbers of subjects** remain the keys to good experimental design.